AUTOMATED SYNTHETIC SCENE GENERATION

DISSERTATION

Ryan N. Givens, Major, USAF

AFIT-ENP-DS-14-S-08

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-DS-14-S-08

AUTOMATED SYNTHETIC SCENE GENERATION

DISSERTATION

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Ryan N. Givens, BS, MSBA, MS

Major, USAF

July 2014

AFIT-ENP-DS-14-S-08

AUTOMATED SYNTHETIC SCENE GENERATION

Ryan N. Givens, BS, MSBA, MS

Major, USAF

Approved:

<u>//signed//</u>	<u>23 Jul 2014</u>
Karl C. Walli, Colonel, USAF, PhD (Chairman)	Date
<u>//signed//</u>	<u>12 Aug 2014</u>
Michael T. Eismann, PhD (Member)	Date
<u>//signed//</u>	<u>15 Aug 2014</u>
Matthew C. Fickus, PhD (Member)	Date

Accepted:

_____   _____

Adedeji B. Badiru   Date
Dean, Graduate School of Engineering
and Management

AFIT-ENP-DS-14-S-08

Abstract

First principles, physics-based models help organizations developing new remote sensing instruments anticipate sensor performance by enabling the ability to create synthetic imagery for proposed sensor before a sensor is built. One of the largest challenges in modeling realistic synthetic imagery, however, is generating the spectrally attributed, three-dimensional scenes on which the models are based in a timely and affordable fashion. Additionally, manual and semi-automated approaches to synthetic scene construction which rely on spectral libraries may not adequately capture the spectral variability of real-world sites especially when the libraries consist of measurements made in other locations or in a lab. This dissertation presents a method to fully automate the generation of synthetic scenes when coincident lidar, Hyperspectral Imagery (HSI), and high-resolution imagery of a real-world site are available. The method, called the Lidar/HSI Direct (LHD) method, greatly reduces the time and manpower needed to generate a synthetic scene while also matching the modeled scene as closely as possible to a real-world site both spatially and spectrally. Furthermore, the LHD method enables the generation of synthetic scenes over sites in which ground access is not available providing the potential for improved military mission planning and increased ability to fuse information from multiple modalities and look angles. The LHD method quickly and accurately generates three-dimensional scenes providing the community with a tool to expand the library of synthetic scenes and therefore expand the potential applications of physics-based synthetic imagery modeling.

Acknowledgments

First, I'd like to thank my research advisor Colonel Karl Walli. Besides having access to your great wealth of knowledge, which made this work possible in the first place, your contagious enthusiasm and optimism pushed me through the difficult stretches of research and writing. I'd also like to thank my committee members Dr. Michael Eismann and Dr. Matthew Fickus for your willingness to serve on the committee and for always being available. Thank you to my parents for training me up in the way I should go. Most of all, thank you to the incredible woman I get to call my wife for turning the challenges of life into our personal adventures.

<div align="right">Ryan N. Givens</div>

Table of Contents

vii

List of Figures

List of Tables

AUTOMATED SYNTHETIC SCENE GENERATION

I. Introduction

1.1   Motivation

The Digital Imaging and Remote Sensing Image Generation (DIRSIG) model is a tool developed and established at the Rochester Institute of Technology over the past decades to generate first-principles-based synthetic images. Before a new remote sensor is built or flown, DIRSIG allows users to model the output of the proposed sensor and thereby anticipate sensor performance. This enables algorithm testing and development, testing of imaging system designs, and the creation of data for training of image analysts allowing processes to be in place prior to the sensor's first flight (Schott, 2007; Schott, et al., 1999; Schott, et al., 2012). Additionally, constructed scenes can be used as a Rosetta Stone to register multiple modalities by registering disparate modalities first to the scene and then backing out the direct registration between the modalities.

To make the output of DIRSIG as realistic as possible, three-dimensional, facetized scenes attributed with spectral information are often built to model real-world locations. While extremely realistic, scenes like DIRSIG's Megascene1, originally built in 2002 and consisting of 1.4 km square tiles of neighborhoods near Rochester, New York, can take months to construct (Ientilucci and Brown, 2003). Physical measurements of homes, trees, and street locations, along with the collection of spectral information using a field spectrometer, are required. Due to the large amount of time and manpower needed to generate these scenes, relatively few are available to the DIRSIG user community.

1

More recently, methods have been developed and significant time savings have been realized using lidar or computer vision techniques to generate the three-dimensional models on which these scenes are based. The Lidar/HSI Direct (LHD) method developed in this research builds on the Lidar Direct method in which a lidar point cloud is facetized using Delaunay triangulation , input directly into DIRSIG in wavefront (.obj) format, then draped with a High-Resolution (HR) image to provide texture information (Walli, 2010). In the Lidar Direct method, the HR imagery is registered to the rasterized lidar return strength using an automated registration process based on the Scale-Invariant Feature Transform (SIFT) algorithm, the RAndom SAmple Consensus (RANSAC) algorithm (Bolles and Fischler, 1981), and a minimization of Root Mean Square Distance Error (RMSDE) of the matched points between the two coordinate systems. Once registered, the HR image acts as texture map for the facetized scene. However, the Lidar Direct method requires human input to attribute spectral information, either from lab-measured spectra or field measurements, to all of the facets in the scene (Walli, 2010).

DIRSIG users are thus confronted with the issue of either being limited to the current DIRSIG library of synthetic scenes or forced to create a scene themselves which better fits the desired modeling parameters. To manually recreate a desert scene with the fidelity of DIRSIG's Megascene1, for example, would take months if similar manual methods were used to capture the scene geometry, texture, and spectral information. The project would require the deployment of an extensive ground team to measure scene geometry and spectra followed by a huge effort to then reconstruct the scene digitally using CAD software to build the geometry and attribute spectral and texture information.

With an automated method, the same user could create a realistic custom scene in a matter of hours if the necessary HR RGB imagery, lidar, and HSI are available. The LHD method presented here provides that automated capability to generate realistic scenes and greatly increase the utility of the DIRSIG model. Additionally, since the LHD method requires only overhead imagery and data to construct a synthetic scene, no ground crew and thus no ground access is necessary for scene construction. This enables the construction of scenes in hostile or otherwise access-limited areas.

1.2 Document Layout

The objective of this research is to demonstrate a method to automate scene generation for use in DIRSIG modeling and test the method on both a modeled dataset, where inputs are known perfectly thus allowing easy comparison of the output results, and a real dataset to test the robustness of the method in a realistic situation. First, relevant work in the field is summarized in Chapter II. Since this method incorporates techniques from several different fields, this section includes discussion on registration methodology, geometry and texture extraction, scene classification and spectral attribution, data fusion, and other scene generation techniques and approaches. The theoretical background is presented next in Chapter III starting with the basics of DIRSIG and the approximations and assumptions that are required for this automated method followed by a discussion of the theory behind the Walli registration method used to relate the input modalities.

Chapter IV first introduces the two datasets used to develop the LHD method. Next, each of the steps of the LHD automated scene generation method are presented

including registration of the disparate input modalities and extraction of the scene geometry and scene texture maps. It then details three approaches to extract the spectral information required for the scene: The material map and material library. Approach 1is an automated method in which the spectral component is extracted using only the HSI input modality. In this case, the material map is limited to the lower resolution of the HSI, relative to the HR and lidar inputs. Approach 2 and Approach 3 attempt to increase the resolution and accuracy of the generated material map by incorporating the spatial information of the higher-resolution modalities into the scene classification. The intended result is a material map with enhanced resolution and thus a higher-fidelity resulting scene. Approach 2 attempts, with limited success, to employ linear unmixing of the HSI pixels and then locate the various component materials in each HSI pixel using the spatial information in the HR imagery and lidar. Approach 2 did not achieve full automation, but is included to show the research progression which results in Approach 3. Approach 3 begins with a high-resolution image segmentation of a fused HR RGB image and rasterized lidar return strength and elevation cube, then uses logic steps to incorporate the lower-resolution spectral information from the HSI to further separate classes in the scene.

An additional goal of this research is to develop a method to evaluate how accurately the modeled scenes represent the original scene. Chapter V presents both qualitative evaluations and quantitative evaluations of scene outputs. The quantitative evaluations are accomplished using a new metric called the Average Band Difference (ABD). It is used to compare synthetic imagery produced using the LHD-generated scenes to the original imagery used to create the scenes. This comparison is done for

4

both a fully-modeled scene in which scene inputs are generated from a high-fidelity

DIRSIG scene and then for a real dataset over a site near Melrose, Florida. The final

chapter, Chapter VI, outlines the major contributions of this research to the modeling and

scene classification. It then presents possibilities for future work and concludes with a

summary of the overall effort.

1.3   Overview of Results

As mentioned above, three approaches are presented for the automated extraction

of the scene spectral component for the LHD method. The three approaches differ only

in their method for generating the material map for the scene. Approach 1 relies only on

the HSI input to generate the material map using the unsupervised Stochastic Expectation

Maximization (SEM) algorithm. This is the first fully-automated approach for extracting

the scene spectral components and thus results in the first full automation of synthetic

scene generation using the LHD method. However, because the HSI is the only input to

determining the material map, the generated material map is limited to the resolution of

the HSI, which is lower than the other two inputs (rasterized lidar and RGB imagery).

Thus, Approach 2 and Approach 3 employ a fused technique to generate the material map

in which the higher-resolution spatial information is incorporated into the generation of

the material map.

The second approach uses the linear mixing model to perform a spectral unmixing

of the HSI pixels with a constrained least squares solver. The resulting material

abundances are used as constraints to then determine the locations of the materials within

each HSI pixel using the spatial information in the RGB and rasterized lidar imagery.

Full automation is not achieved which Approach 2 due to the inability to satisfactorily automate the selection of material endmembers with which to perform the linear unmixing of the HSI pixels. Endmembers are therefore selected by hand to allow an evaluation of the resulting HR material map using this approach to determine whether additional efforts should be spent in automating the endmember selection process. As is shown in Chapter III, the resulting material maps using this method, even with hand-selected endmembers, are not of sufficient quality to invest additional time in this approach.

Approach 3 builds on Approach 2 but runs in reverse sequence by first using a clustering method to segment the higher-resolution inputs (a fused RGB and rasterized lidar cube) using the Iterative Self-Organizing Data analysis technique (ISODATA). It then uses the registration transformation matrices to reference the HSI spectral content to separate segmented areas consisting of different materials using the Spectral Angle Mapper (SAM) technique. Full automation is achieved using this approach in addition to generating a HR material map.

The two fully automated methods (Approach 1 and Approach 3) are used to generate scenes using input data from both a modeled and real-world site to evaluate the quality of the recreated scenes. To perform this evaluation DIRSIG-generated synthetic RGB imagery and HSI is compared to the input RGB imagery and HSI. The comparison metric used is explained in detail in Chapter V, but is basically an average difference per band at each pixel which is then averaged across the scene. Table 1 summarizes the three approaches used and the average band difference (ABD) results for Approach 1 (the HSI-only approach) and Approach 3 (the fused segmentation approach) when used to

reconstruct the modeled VanLare scene. As the table shows, the HSI-only and

Segmentation approaches both produce relatively accurate reconstructions of the original

VanLare scene with respect to the average per pixel and per band difference in RGB

intensity (on a scale of 0 to 255) and HSI reflectance (on a scale of 0% to 100%). While

the ABD results do not show a large difference between Approach 1 and Approach 3,

Chapter V presents an alternative metric to compare the approaches over smaller

windowed regions.

Table 1. Summary of the three approaches investigated to automate the generation of a material map for the LHD method (NMM and LMM stand for Normal and Linear Mixture Model, respectively).

Approach	Fused/ Sharpened	Fully Auto- mated	Spectral Data Model Used	Primary Algorithm Used	RGB ABD (Digital Counts)		HSI ABD (% Reflect)	
					Avg	Std	Avg	Std
1: HSI-only	No	Yes	NMM	SEM	4.5	6.5	2.0	2.0
2: Unmixing	Yes	No	LMM	Constrained Least Squares	-	-	-	-
3: Segmentation	Yes	Yes	NMM	ISODATA/ SAM	4.5	6.2	2.2	2.3

The alternative metric uses a threshold percent difference between the input

imagery and recreated imagery to determine the number of pixels within 15% of the

original input imagery. While presented as a topic for future research and improvement,

the metric does show that Approach 3 results in improved accuracy in regions with fine

spatial details. Thus, determining which approach to use depends on the GSD of the

remote sensor to be modeled. The HSI-only approach may be sufficient if the user is

attempting to model a sensor with a larger GSD than the input HSI. However, if the GSD

of the sensor to be modeled is smaller than the input HSI, the quality of fine details in the

synthetic scene are noticeably improved when Approach 3 is used.

II. Literature Review

2.1   Chapter Overview

Developing methods to automate the generation of three-dimensional scenes for various types modeling has become more prevalent in recent years for uses such as city planning (Föerstner, 1999), environmental modeling (Brenner, 1999), navigation (Auer, et al., 2010; Brenner, 2005), games and entertainment (Hearn and Baker, 1997), military planning (Mason, 2004) and, for the focus of DIRSIG modeling, remote sensing design and applications (Gurram, et al., 2007; Ientilucci and Brown, 2003; Lach, et al., 2009; Schott, 2007; Schott, et al., 2012; Walli, 2010). In many cases, as in the modeling of wind patterns or wave propagation through city skylines, three-dimensional models may be required only to be accurate in their three-dimensional geometry. In uses such as city planning, games and entertainment, and military planning, scenes have the additional requirement to be accurately attributed with visible color and texture information as well. The amount of research published on autonomously-generated, fully-textured and spectrally accurate (beyond the visible wavelengths) three-dimensional models for the use of modeling remote sensing instruments is relatively small.

This chapter summarizes previous work in the various disciplines involved in generating a textured and spectrally attributed three-dimensional scene. Since no single sensor currently captures all of the information required for these models, the first discipline involved is multimodal registration. Next, current techniques in geometry extraction and texture attribution are summarized. Spectral classification using HSI alone and HSI fusion techniques are then presented. Finally, previous efforts which explore the

automation process of bringing together the various inputs to generate the fully attributed scene are summarized.

2.2   Multimodal Registration Techniques

Whether accomplished manually or using an automated approach, registration is a first step to accomplish a plethora of tasks in the fields of remote sensing, medical imaging, and computer vision. Various methods including the registration of images of different viewpoints, times, modalities, and real imagery to synthetic imagery have been specifically developed for each of these cases. While a full investigation of all of the registration approaches is beyond the scope of this document, an overview of the different types is included here to show why the SIFT, RANSAC, RMSDE minimization approach developed by Walli is an appropriate method for the registration of the various modalities involved in automated scene construction.

The survey of image registration methods by Zitova and Flusser published in 2003 provides a good overview of the types of techniques and steps involved in registration. Most image registration methods fall into one of two types of approaches: area-based or feature-based. Using one of these two types of approaches, registration methods generally follow four steps in accomplishing the registration: feature detection, feature matching, transform model estimation, and image resampling and transformation (Zitová and Flusser, 2003).

For the feature detection step of the registration process, area-based methods depend on matching intensity patterns in one image to intensity patterns in another image using some type of correlation metric. While this can produce successful results when

the two images have similar GSD, illumination conditions, and viewing angles, correlation techniques may struggle when trying to register images from different types of modalities where these image parameters are likely not similar between the modalities (Walli, 2010).

Feature based registration methods on the other hand, tend to use a higher level of information than image intensity to detect features in images. This is useful because the descriptors can be developed in such a way as to be invariant to common differences between images such as scale and rotation. The SIFT algorithm is often used because the keypoint descriptors are not only invariant to scale and rotation, but also robust across a range of affine distortions, changes in viewpoint, noise, and illumination (Lowe, 2004). This is ideal for the multimodal registration required for scene extraction because the HR imagery, lidar, and HSI sensors are very different and even when all three are available over a scene; they are likely collected at different illuminations and many times at different times of the year. SIFT is also efficient in the way it matches keypoints between images using a simple distance metric.

SIFT, like many registration techniques can be combined with other culling or refinement methods to improve registration. Hasan, et al., combined SIFT with a technique to also consider spatial relationships of keypoints to improve the reliability of selected matches when registering multi-spectral images (Hasan, et al., 2010). Another example of SIFT application in registering remote sensing images is published by Song, et al. This research improves the robustness of SIFT by employing both a spatial analysis of matched points along with a mutual information metric generated using Lissajous curves (Song, et al., 2010). As a final example, Walli found SIFT to be particularly

11

robust in its ability to register imagery between modalities including scene to model registration (Walli, 2010). While these are just a few examples, literally thousands of papers (24,321 according to Google Scholar at the time of this writing) have been published on various registration techniques using the SIFT algorithm leading to its status as the gold standard for the registration of remote sensing imagery.

RANSAC is another widely used algorithm in the area of registration research due to its ability to quickly sort through potential matches and eliminate incorrect matches (Walli, 2010). Using RANSAC to cull incorrectly identified matches is therefore a common second step to the SIFT algorithm. RANSAC can be used to enforce appropriate relationships on the set of matches using epipolar geometry and homography constraints (Hasan, et al., 2010; Walli, 2010; Wong and Clausi, 2007). This leads naturally into the next step of the registration process of estimating the transform model.

To estimate the transform model, the set of matched points are used to extract the transformation which relates the pair of images in the registration process. Generally, as is the case in imagery used in the LHD method, some assumption about the transformation relation can be made. When dealing with georeferenced imagery taken from a nadir viewing geometry, the transform can be assumed to be affine if not more restricted to a linear conformal transformation. With this assumption, only three matched points are required to solve for the six unknowns in the affine transformation matrix and thus when more than three matches are available, the problem is over-determined and solving for the transformation is accomplished by determining the transformation which results in the lowest amount of error for the set of match points (Zitova and Flusser,

2003). In this research, MATLAB's cp2tform function is used to performed this

operation (MATLAB, 2010).

The final step in any registration process is to choose the type of interpolation

used to map one of the images to the coordinate system of the other image. For the

registration in this work, MATLAB is used to perform the transformation using the

imtransform function and its default bilinear transformation is selected when

transforming all imagery except the HSI. In the case of HSI, it is imperative that the

spectral component is not polluted by incorporating local spatial information in an

interpolation. Thus, nearest neighbor interpolation is used when transforming HSI

imagery to another coordinate system.

2.3   Geometry and Texture Extraction

The geometry and texture for a scene can be generated manually by transferring

physical measurements from a site into a computer assisted drawing program and then

attributing each facet with appropriate texture and spectral information. The result,

depending on the amount of time invested in making measurements and drawing detailed

geometries, can be an extremely accurate and detailed scene as in DIRSIG's megascene1

(Ientilucci and Brown, 2003). However, the focus for this work is to reduce the time and

cost involved in making these accurate wide-area scenes, thus semi-automated and

automated techniques are reviewed here.

Semi-automated and automated geometry extraction is typically accomplished

using photogrammetric techniques or lidar-based techniques. Each set of techniques has

its advantages and disadvantages. Photogrammetric techniques rely on the ability to

determine the scene geometry using images taken from multiple viewpoints of a scene. If available, multiple images can be registered and related by finding match points and extracting the fundamental matrices defining the epipolar geometry of the various camera positions. The scene geometry is then extracted from this information (Cheng and Gong, 2006; Hartley and Zisserman, 2003; Föerstner, 1999). A strength of this technique is that, once the geometry is known, texture from the registered images is easily applied to the geometry to create realistic looking scenes. A drawback of this technique is that many images may be needed to generate the dense point cloud required to construct the geometry. Lidar-based techniques have an advantage in that the scene geometry is measured directly by the instrument and no additional processing is therefore required to generate the geometry besides facetization of the lidar point cloud. Texture information for non-vertical surfaces in the scene can be draped over the geometry using a rasterization of the lidar return strength or by registering and draping a separate aerial image over the geometry (Brenner, 2005; Lach, et al., 2009; Schenk and Csathó, 2002; Schott, et al., 1995; Walli, 2010). Additional off-nadir imagery is required, however, to generate textures for vertical facets and color imagery is necessary to generate the proper reflectance values if the scene is to model color images.

Whether photogrammetric or lidar techniques are used, the generated geometries can be refined and the number of facets reduced for man-made structures in the scene by extracting building footprints and restricting items like building walls and surfaces to be single, flat facets (Khoshelham, et al., 2005; Lach, et al., 2009). Additionally, methods have been developed which have the ability to extract tree center locations from the geometry point clouds and then place a higher fidelity representation of a tree in those

locations (Ientilucci and Brown, 2003; Lach, et al., 2009; Li, et al., 2012; Lin, et al., 2011; Wu, et al., 2013; Zhang and Qiu, 2012). These refining techniques could be incorporated into the LHD methodology presented here, but, as the purpose of this work is to generate wide-area scenes and thus model remote sensing instruments with at least 1 meter GSD, this level of fidelity was deemed unnecessary for the initial development of the LHD method.

2.4   Scene Classification using HSI and Fused Methods

Automating the generation of and then refining the spectral component of the scene building process is the primary focus of the research presented in this dissertation and the final step required to fully automate scene generation from coincident HR imagery, lidar, and HSI. A vast amount of research exists for the segmentation and classification of all three of these input datasets involved in the LHD method.

2.4.1   Hyperspectral-Only Classification

As is detailed in Chapter IV, scene classification in Approach 1 is accomplished using a standard HSI-only classification algorithm called Stochastic Expectation Maximization (SEM) (Celeux and Diesbolt, 1986; Masson and Pieczynski, 1993). This algorithm is included in the approach due to its popularity as an unsupervised classification method as well as its requirement that only a single initial input of initial class number is needed to start the algorithm (Eismann, 2012; Givens, et al., 2012). Similar in its ubiquity to SIFT and RANSAC, SEM is another example of a widely used algorithm in its respective community of scene classification. Originally developed by Celeux and Diesbolt in 1986, SEM performs unsupervised segmentation of data using

maximum likelihood clustering based on posterior probabilities (Masson and Pieczynski, 1993). This is equivalent to assigning spectra to classes based on Mahalanobis distance as in a simple quadratic clustering method, but in SEM classes are allowed to overlap and thus some randomness is introduced (Eismann, 2012). Example resulting material maps for both modeled and real HSI are presented in Chapter VI.

2.4.2   Fusion Classification Approach Using the Linear Mixing Model

While the unsupervised SEM segmentation method of Approach 1allows for the full automation of the generation of the material map and spectral library for the LHD method, this research also presents the application of fusion-based techniques to improve the resolution and accuracy of the material map. The first such approach, Approach 2, employs the linear mixing model to constrain and unmix the HSI pixels resulting in a higher-resolution material map. The linear mixing model is a simple and physical model for HSI in that it assumes each pixel in a hyperspectral image could consist of multiple material types and that each pixel's spectrum is thus a linear combination of the spectra of the pure materials captured in the pixel. The coefficients in the linear combination therefore refer to the percentage (abundance) of each material which exists in the pixel. Thus, if the endmembers (the pure spectra for each of the materials in the scene) are known, each pixel is simply a linear combination of these endmember spectra and all of the coefficients are restricted to be both positive and sum to one (Eismann, 2012; Gross and Schott, 1996). While seemingly simple and intuitive, perfect endmember spectra are not easily determined or known. Thus, the nonnegativity and additivity constraints can be adjusted based on a user's confidence in their actual knowledge of the endmembers (Eismann, 2012).

16

Various approaches are available to attempt to determine the endmembers that exist within any given hyperspectral scene. Pixel Purity Index (PPI) (Chang and Plaza, 2006) and N-FINDR (Winter, 1999) are two such approaches which are employed with limited success in Approach 2 of Chapter IV. Both of these methods attempt to find endmembers based on their extreme locations within the multidimensional point clouds which make up a hyperspectral dataset. PPI randomly generates lines on which to project the points in the hyperspectral dataset. By repeatedly generating these lines and calculating the projections of the points in the point cloud, PPI determines which points in the hyperspectral dataset routinely result in large projections onto the randomly generated lines. These points are assumed to be pure pixels and thus endmember candidates. N-FINDR uses a similar logic with a volume-based approach. Based on a user input of number of pure materials in the scene, N-FINDR computes the multidimensional volume for the points in the data cloud generally starting with a random set then using a gradient metric to continually select points which result in larger volumes. The set of points which results in the largest volume is then assumed to be composed of endmember candidates (Eismann, 2012).

Once endmember candidates are determined, a least squares solver can be used to find the best fit linear combination of endmember abundances which make up each pixel in a hyperspectral image. While intuitive, the linear mixing model is limited in how well it actually fits the data. First, mixing may not be completely linear between materials. Also, the single endmember model is generally not able to represent a real class in a scene. A pure grass pixel, for example, is generally composed of various species of grass as well as grass with different amounts of water, health, fertilizer, etc. all of which have

an effect on the reflectivity of grass. The linear mixing model also struggles with representing low lighting, or shade, pixels. A general work-around to the shade case is to have a shade class which is mixed into shade pixels, but a shade class is not desirable for a modeled spectral library in which one hopes to create a scene which will generate accurately synthetic scenery regardless of illumination angle (Eismann, 2012). Finally, determining the number of desired endmembers is also not a trivial task. Generally this is done through some sort of dimensionality reduction where the number of pure materials in a scene is inferred from the number of principle component bands determined to be significant. However, determining the number of significant principle component bands can be somewhat arbitrary if based on some threshold of variance captured. Alternatively, the number of significant principle components can be estimated using more complex estimations of noise in the data (Chang and Du, 2004; Eismann, 2012).

As is shown in the presentation of Approach 2 in Chapter IV, the linear mixing model is used to determine material abundances in hyperspectral pixels, which are then transformed to the HR and lidar GSDs in an attempt to locate the various materials within each pixel according to the unmixed abundances. However, this approach runs into challenges first in finding an algorithm to automate the selection of endmembers then, even with hand selected endmembers, in satisfactorily determining abundances and locating material positions with the hyperspectral pixels.

2.4.3 Fusion Approach using ISODATA and Rules-based Algorithm

The final and most successful fusion based approach to scene classification found in this research, Approach 3, begins with a fusion of the HR information from the RGB imagery and lidar data. This portion of the method is based on the work of Lee and Shan

published in 2002 in which they improved coastal coastline mapping using ISODATA to classify a fused cube consisting of multispectral imagery and lidar data (Lee and Shan, 2002). In this research, as is shown in Chapter IV, the HR imagery is fused with rasterized lidar elevation maps and return strengths instead of multispectral imagery. Thus, both components are of a relatively high resolution compared to the HSI GSD. Once the two HR components are used to generate a HR material map, class confusion is resolved by performing a spectral comparison of pure connected component regions in the HSI then applying logic-based rules.

2.4.5   Other Fusion-Based Approaches

While the three approaches outlined above are the methods applied in this research, there are a multitude of alternative approaches to generating material maps both from HSI alone and HSI fused with other modalities. One popular approach is the sharpening of the HSI which incorporates the high frequency spatial information from RGB or panchromatic imagery into HSI imagery to increase the spatial resolution of the HSI. Generally the HR spatial information is incorporated using a domain change to the frequency regime with a Fourier or Wavelet transform, or a transformation to principal component space. While many of these methods result in imagery that is sharper to the human eye, the radiometry of the original HSI is harder to preserve and often bleeding and interpolation artifacts show up in the resulting imagery (Borel, et al., 2010; Eismann and Hardie, 2005; Svab and Ostir, 2006; Walli, 2003). Even with these drawbacks, these sharpening techniques could have been used in this work, but the success of Approach 3 made investigation into these techniques unnecessary. Additionally, Approach 3 has a

19

relatively small computational cost since it uses a simple distance calculation in the application of the ISODATA algorithm on the five bands of the fused RGB/lidar cube.

Other techniques for fusing lidar and HSI information exist such as extracting surface roughness from lidar to compare with HSI classification results (Mundt, et al., 2006; West and Resmini, 2009) and an increasing number of sensors now exist to capture the complementary information of lidar and HSI (Asner, et al., 2007). Other works incorporate lidar height information to perform automated target detection. In this case, targets are assumed to be above ground and of a certain size allowing for potential target areas to be identified in lidar elevation maps before comparing spectra in HSI (Kanaev and Walls, 2011). Other research efforts focus on the identification of tree types and locations using the lidar elevation information from the lidar fused with spectral information gained from HSI (Daplonte, et al., 2009). The wide range of available fusion techniques and algorithms leaves a nearly inexhaustible supply of methods to best classify scenes using hyperspectral, HR, and lidar fusion.

2.5 Generating Spectrally and Texturally Attributed Three-dimensional Scenes

Two previous research efforts focused on the semi- and full automation of scene generation for use in DIRSIG with full spectral attribution of the visible through short-wave infrared wavelengths. In the first, Lach et al. did a great deal of work focused on the high-fidelity extraction of the scene geometry from lidar, then applied texture and spectral information from RGB imagery and HSI. Many portions of the texturing and spectral attribution are automated in his research, but the registration between the lidar and HSI is accomplished by hand-selecting matching ground control points.

Additionally, extraction of the material map and spectra from the HSI are accomplished using a supervised classification routine with manual refinement of the material map (Lach, et al., 2009). In the second research effort, Walli developed a robust registration method to automate the registration of the geometry and texture components and fully automated the geometry extraction and draping of texture maps in a method dubbed the Lidar Direct method. However, spectral attribution is accomplished manually using a combination of supervised classification of the HR imagery and hand-editing, ultimately linking an existing spectral library to the geometry and texture (Walli, 2010). The method presented in this paper, the Lidar/HSI Direct method (LHD), extends the Lach and Walli methods to fully automate the spectral attribution and spectral library extraction using the robust registration technique developed by Walli and an unsupervised classification of the HSI to create the material map and pull reflectance spectra directly from the HSI. Additionally, Approach 3 employs fusion techniques to incorporate the higher spatial information available in the HR imagery and lidar return to improve the resolution and accuracy of the HSI material map. The result is a fully automated approach to generate accurate, three-dimensional scenes with both texture and spectral attribution for use in the DIRSIG modeling software.

III. Theoretical Background

This chapter presents the theory behind the physics-based modeling used in DIRSIG especially in how it relates to the use of these autonomously generated scenes. Next, the robust registration method used to spatially relate the three input modalities is presented. The Walli registration process uses three primary algorithms: SIFT, RANSAC, and average RMSDE minimization. The details of these three algorithms are presented along with an explanation of how they are included in the registration process. The automated registration between the three input modalities of HR, lidar, and HSI is an essential first step to enabling the automated extraction of the synthetic scenes presented in Chapter IV.

3.1   DIRSIG Physics-Based Modeling Overview

DIRSIG has been and continues to be developed at Rochester Institute of Technology's Chester F. Carlson Center for Imaging Science. It is an extremely powerful modeling package providing the tools to generate synthetic imagery for both passive and active sensors from 0.4 μm to 20 μm and beyond with recent work extending the DIRSIG model into the longer synthetic aperture radar wavelength regime. The DIRSIG model is able to handle both reflective and emissive properties of materials using measured emissivities or, if measured emissivities are not available, thermal modeling for mid and long-wave infrared regions. This research focuses on the visible through near infrared and short wave infrared region (VNIR/SWIR) of the electromagnetic spectrum (0.4 μm to 2.5 μm) because of the availability of HSI datasets in this range. Thus, the theoretical discussion presented here focuses on reflective

properties of materials. However, in cases where mid-wave or long-wave infrared HSI is available, the LHD method could be similarly applied to generate synthetic scenes in those regions as well.

DIRSIG outputs are useful for a variety of purposes including sensor prototyping, algorithm testing and training, and analyst training. The ultimate goal of DIRSIG is to provide "an integrated environment for image system simulation" in which users are able to "prototype, evaluate, and demonstrate the capabilities of next generation imaging systems." (DIRSIG Course Lecture Materials, 2011)

DIRSIG is a physics-based ray tracing modeling software that allows a user to specify all scene, atmospheric, sensor, and illumination conditions which are then used to generate a synthetic image at the defined sensor based on the input conditions. At its most basic level, DIRSIG relies on the radiation propagation equation

$$L_p(\lambda) = \tau_a(\lambda)L_u(\lambda) + L_a(\lambda), \tag{3.1}$$

where $L_p(\lambda)$ is the spectral radiance at the aperture of the sensor, $\tau(\lambda)$ is the atmospheric path transmission, $L_u(\lambda)$ is the upwelling radiance, and $L_a(\lambda)$ is the path radiance and all of these terms are dependent on the wavelength, λ. The upwelling term from Equation (3.1), $L_u(\lambda)$, for a solid surface consists of two pieces, emitted and reflected radiance such that

$$L_u(\lambda) = L_e(\theta_r, \phi_r, \lambda) + L_r(\theta_r, \phi_r, \lambda), \tag{3.2}$$

where L_e and L_r are the emitted and reflected radiance depending on the facet's angular orientation to the sensor (defined by θ_r and ϕ_r) and wavelength λ. The emitted and reflected components from Equation (3.2) are then further defined as

$$L_e(\lambda) = B(\lambda, T)\left[1 - \int_0^{2\pi}\int_0^{\frac{\pi}{2}}\rho_{BRDF}(\theta_r, \phi_r, \theta, \phi, \lambda)\cos\theta\sin\theta\, d\theta d\phi\right], \qquad (3.3)$$

and

$$L_r(\lambda) = \rho_{BRDF}(\theta_r, \phi_r, \theta, \phi, \lambda)E_s(\theta_s, \phi_s, \lambda)$$

$$+ \int_0^{2\pi}\int_0^{\pi/2}\rho_{BRDF}(\theta_r, \phi_r, \theta, \phi, \lambda)L_d(\theta, \phi, \lambda)\cos\theta\sin\theta\, d\theta d\phi, \qquad (3.4)$$

where $B(\lambda,T)$ is the blackbody spectral radiance distribution for the facet at temperature

T, ρ_{BRDF} is the bidirectional reflectance distribution function (BRDF) which is a material

property relating the angle of incident irradiance to the angle of reflected radiance, $E_s(\theta_s,$

$\phi_s, \lambda_s)$ is the solar irradiance, $L_d(\theta, \phi, \lambda)$ is the downwelling radiance, and the various

angles are defined as in Figure 1 below.
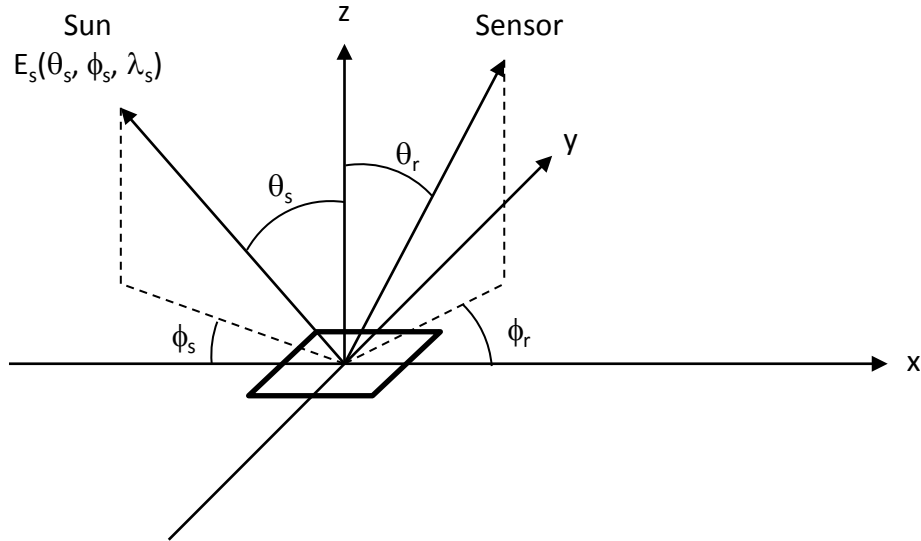
Figure 1. Geometry for a single facet (Eismann, 2012).

Using the Beard-Maxwell BRDF model, the BRDF from Equations (3.3) and (3.4) is

composed of specular, diffuse, and volumetric terms such that

24

$$\rho_{BRDF}(\theta_r, \phi_r, \theta, \phi, \lambda)$$
$$= \rho_s(\theta_r, \phi_r, \theta, \phi, \lambda) + \rho_d + \frac{\rho_v(\lambda)}{\cos\theta + \cos\theta_r}. \tag{3.5}$$

If facets are approximated as diffuse Lambertian reflectors, as in this research, the BRDF

specular and volumetric terms (ρ_s and ρ_v) from Equation (3.5) are assumed to be zero and

therefore loses its angular dependence such that

$$\rho_{BRDF}(\theta_r, \phi_r, \theta, \phi, \lambda) = \rho_d = \frac{\rho_{HDR}(\lambda)}{\pi} = \frac{\rho_{DHR}(\lambda)}{\pi}, \tag{3.6}$$

where ρ_{HDR} is the hemispherical directional reflectance which characterizes the reflected

irradiance in a particular direction if the facet were illuminated by a perfectly diffuse

source and ρ_{DHR} is the directional hemispherical reflectance which is the ratio of the total

reflected irradiance to the directional incident irradiance. Using the Lambertian

simplification of Equation (3.6) and recognizing that thermal emission is dominated by

solar irradiance in the VNIR/SWIR region (i.e. $L_e \cong 0$), Equation (3.1) simplifies to

$$L_p(\lambda) = \frac{\tau_a(\lambda)\rho_{DHR}}{\pi}[E_s(\theta_s, \phi_s, \lambda) + E_d(\lambda)] + L_a(\lambda), \tag{3.7}$$

where $E_d(\lambda)$ can be approximated by

$$E_d(\lambda) = \Omega_d L_d(\lambda), \tag{3.8}$$

if the angular dependence of the downwelling irradiance is ignored and Ω_d is the solid

angular region of the unobstructed sky. While DIRSIG can handle specular facets,

Equation (3.7) describes the radiative transfer equation DIRSIG uses (along with

MODTRAN to calculate the path transmission, solar irradiance, downwelling and path

radiance terms) to trace rays through a scene, off diffuse facets, and back up to the sensor

to calculate at-aperture radiance.

DIRSIG handles interaction with the atmosphere, including both transmission and scattering effects, using MODTRAN. MODTRAN has been under development by the Air Force Research Laboratory since 1987 and is a well verified and common tool for atmospheric propagation modeling for the community (Berk, et al., 1998). DIRSIG modeling incorporates both reflected and emitted radiation as well as multiple bounce interactions. For example, a target in a scene may be illuminated directly by the main source in the scene, but may also receive reflected illumination from nearby objects.

DIRSIG also includes the capability to model bidirectional reflectance distribution functions for materials that are included in a scene. For the purposes of this work, however, all materials in the scene were assumed to be Lambertian materials. For BRDF to be included in the material properties, identified materials would need to be matched to laboratory measured BRDFs. Including a BRDF for the scene materials is beyond the scope of this work, but could be an area for future research.

Materials in DIRSIG are defined spectrally using their surface emissivity as a function of wavelength. Another simplification is used here in that all materials in the autonomously generated scenes are assumed to be opaque and at thermal equilibrium. Using these assumptions and conservation of energy, material transmissions are zero and material emissivities are equal to material absorption. Kirchoff's law thus allows the determination of emissivity from reflectance, the value obtained from atmospherically compensated HSI, by simply subtracting reflectance from one as shown in Equation (3.9) below:

$$Emissivity = 1 - reflectance. \qquad\qquad (3.9)$$

Thermodynamic properties of materials can be included in DIRSIG, but are also ignored for this work because these properties are not directly measured by the three modalities used here and are negligible effects in the VNIR/SWIR region considered in this work. This limits the scenes modeled using the automated technique to visible and shortwave modeling only when VNIR/SWIR HSI is used. For models to accurately produce imagery in the mid-wave or long-wave regimes, the spectral signatures of the materials in the scene as measured by the HSI would need to be matched to lab measured library materials that include the bulk properties necessary for accurate thermodynamic modeling beyond the short-wave regime.

Geometries in DIRSIG generally consist of three-dimensional facetized vertices. Multiple geometries are easily incorporated in a scene by simply specifying the insertion coordinate for a new object. This is helpful for inserting a user-defined geometry for a target, like a plane or vehicle for instance, into an existing scene. For this work, the autonomously generated scenes consist of a faceted ground and building geometry with a separate geometry to handle tree canopies. In the LHD method, the ground and building vertices in the generated scenes are tinned using a Delaunay triangulation facetization. Since material transmissions are defined to be zero, it is important to handle the trees in a different manner so that transmission through tree canopies can occur. Thus, instead of tinning vertices identified as tree vertices, a 0.5 m by 0.5 m tile is placed at each tree return location. These floating tiles, while still a somewhat crude model of a tree, at least allow light to pass through the tree canopies so that they do not create hard shadows. DIRSIG can also handle voxelized geometry and mathematical geometries (such as spheres and cubes), but these are not currently used in the LHD method.

The modeling of imaging instruments in DIRSIG is very flexible. Most commonly, DIRSIG is used to provide the "at sensor" integrated radiance and the actual imaging system optics are not included. Thus, the sensor is defined by focal length, the pixel array geometry, and spectral responsivity. Spectral responsivity can be defined using basic functional shapes like Gaussian or triangular centered on a wavelength, or can be completely manually defined using a modeled or measured spectral responsivity.

The final steps in generating synthetic imagery with DIRSIG are to specify the sensor location over the scene, including any motion if the sensor captures continuously like a push broom or whisk sensor, and time of the collect. The user can also specify multiple truth output files like scene geometry or material map to gain access to what is often the most useful feature and purpose of any such modeling: Perfect ground truth.

3.2   DIRSIG Scene Components

DIRSIG scenes consist of three-dimensional, facetized surfaces attributed with spectral and texture information. The scenes can be generated in a number of ways and to varying degrees of fidelity in both their spatial and spectral resolution, as well as their thermodynamic properties. However, this research deals only with visible through short-wave infrared wavelengths, so thermodynamic properties are ignored.

Scenes like DIRSIG's MicroScene2, a scene which is available for download to DIRSIG users and models Camp Eastman in Rochester, NY, consists of carefully built trees that are placed within the scene to accurately match the real trees at the modeled site. The creators of this scene used specialized tree growing software which allows the trees to be modeled down to the individual leaves and pine needles. Scenes with this

level of fidelity are therefore very good for HR simulations, but the time involved in generating them and the amount of memory needed to model to such a high resolution limits the size of the scene.

For large area scenes, DIRSIG users have a few options available for download. One of those options is MegaScene1, which was built around 2002 and models Irondequoit, NY, a suburb of Rochester, NY. This scene consists of a general ground geometry which is classified using an overlaid material map and includes texture using an overlaid texture map. Geometry above the ground level, representative trees and buildings, are individually constructed and planted in the scene to match the actual locations, if not the exact building or tree type, of the real structures and trees. This scene currently has approximately one meter resolution with a massive rebuild underway using the CityEngine software package. Inputs for this scene therefore include the ground-geometry digital elevation map, some aerial imagery which provides the texture map and material map, spectral measurements made using portable field spectrometers, and geometric measurements for placing buildings and trees within the scene. Construction of a medium fidelity scene like this takes months and thousands of man hours to collect the field measurements and then actually build the scene. As a result, there are few of these large area scenes available for the DIRSIG user community. Thus, DIRSIG is limited in its ability to model sensors with wide fields of view, not because it lacks the capability to perform the modeling, but because the few available large scenes may not be representative of the environment in which the user intends to deploy the sensor.

To address this problem, this research proposes to automate the entire process of generating large area scenes. Fortunately, with the increased collection of HR imagery,

hyperspectral imagery, and lidar, the necessary components to build scenes of many real world sites are available. Automated scene generation is therefore possible by gathering the datasets, fusing them, and extracting the necessary geometry, texture, and spectral information in the format DIRSIG requires for a scene.

3.3 Registration

Before the scene inputs needed for DIRSIG can be extracted from the input modalities, a high-confidence and robust registration method must be used to register the modalities. The Walli registration method used in this research is based on a registration technique which finds potential match points using SIFT, refines the set of matched points using two iterations of the RANSAC algorithm where epipolar and homogeneous constraints are enforced, and then settles on a final set of matched points by recursively eliminating high-error matches until the average RMSDE for the entire set falls below a user threshold (Walli, 2010). The resulting affine transformation matrices relating the three modalities allow the extracted DIRSIG inputs to be properly related and overlaid so that the scene can be accurately modeled.

3.3.1 Image Preparation for Registration

Before SIFT is used to detect and match keypoints, the input modalities must be prepared for registration. First, the higher resolution inputs are resampled to the HSI Ground Sample Distance (GSD) using a bilinear interpolation. The RGB image is converted to grayscale by "eliminating the hue and saturation information while retaining the luminance" using MATLAB's rgb2gray function (MATLAB, 2010). Representative spectral bands are then selected from the HSI (~450 nm, ~550 nm, and ~650 nm for the

30

RGB image and ~1060 nm for the rasterized lidar return strength). A weighted selection

of multiple HSI bands could be used to better approximate the light gathered by the HR

bands, but selecting just the narrow center-wavelength RGB bands from the HSI proved

to be sufficient for registration. Next, the red, green, and blue center-wavelength HSI

bands are converted to grayscale using the same process as on the HR imagery. Finally

all inputs are histogram equalized to a flat histogram using MATLAB's histeq function

(MATLAB, 2010). A flowchart of the image preparation process is shown below in

Figure 2.
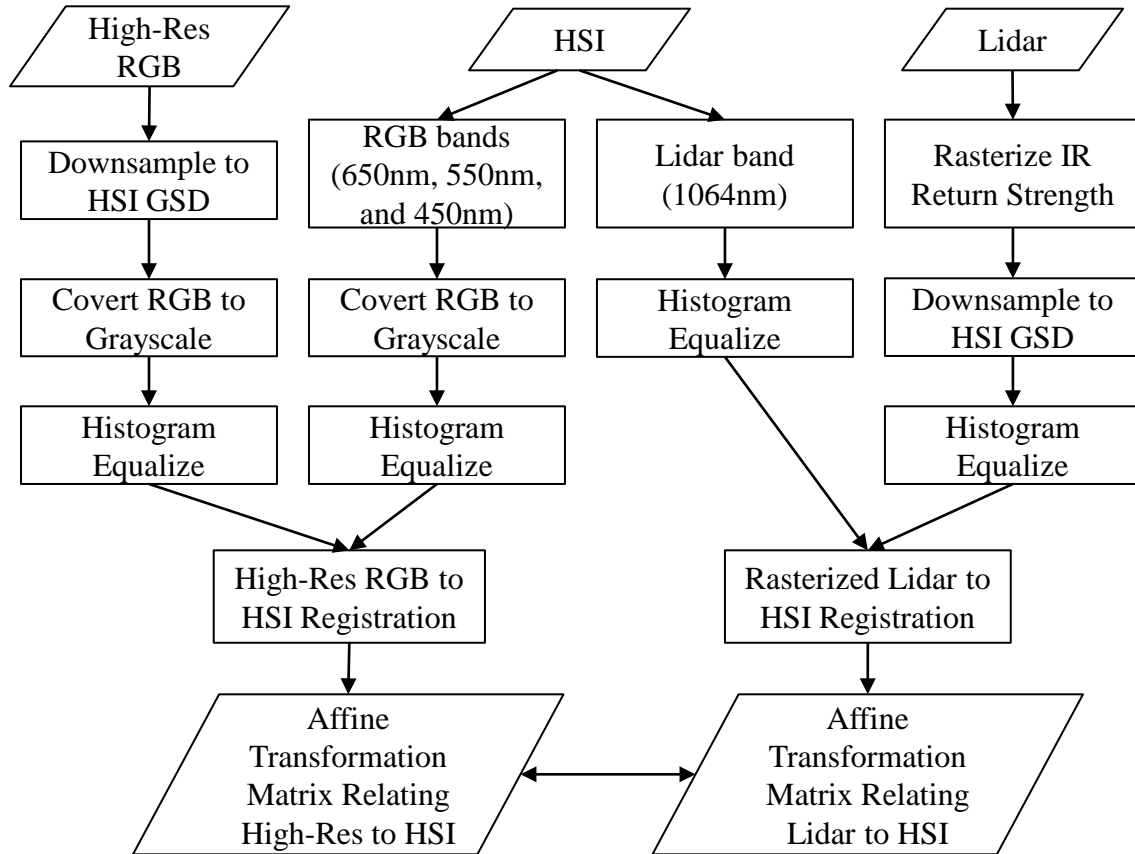
Figure 2. Flowchart of data preparation before registration process leading to the output affine transformations relating the three input modalities.

3.3.2   Using SIFT to Determine Potential Matches

After the input imagery is appropriately prepared, the registration method begins

with the SIFT, or Scale-Invariant Feature Transform, algorithm. The SIFT algorithm is

an edge detection algorithm developed by David Lowe (Lowe, 2004). The algorithm

accepts two grayscale images as input and then identifies invariant keypoints in each

image (usually on the order of 3000-5000 for a 300 by 500 pixel image). To identify

keypoints in an image, the SIFT algorithm first increasingly blurs the image using a

Gaussian kernel as shown in Equation (3.10)

$$L(x, y, \sigma) = \ G(x, y, \sigma) * I(x, y), \tag{3.10}$$

where L is the output blurred image, $*$ is the convolution operator, σ is the scale factor, I

is the input image, x and y are the coordinates in image space, and the Gaussian kernel

from Equation (3.10), G(x,y,σ) is defined as

$$G(x, y, \sigma) = \ \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{3.11}$$

Each blurred image, L is then subtracted from the next less-blurred image to create a

range of difference of Gaussian images:

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y)$$
$$= L(x, y, k\sigma) - L(x, y, \sigma). \tag{3.12}$$

The original image is then degraded to half of its original resolution and the process is

repeated to create the next octave. This process is shown pictorially in Figure 3 below.

Each pixel in each difference of Gaussian image is then compared to the eight pixels

surrounding it as well as the nine pixels above and nine pixels below in the stack of

difference of Gaussian images to identify local maxima or minima as shown in Figure 4.
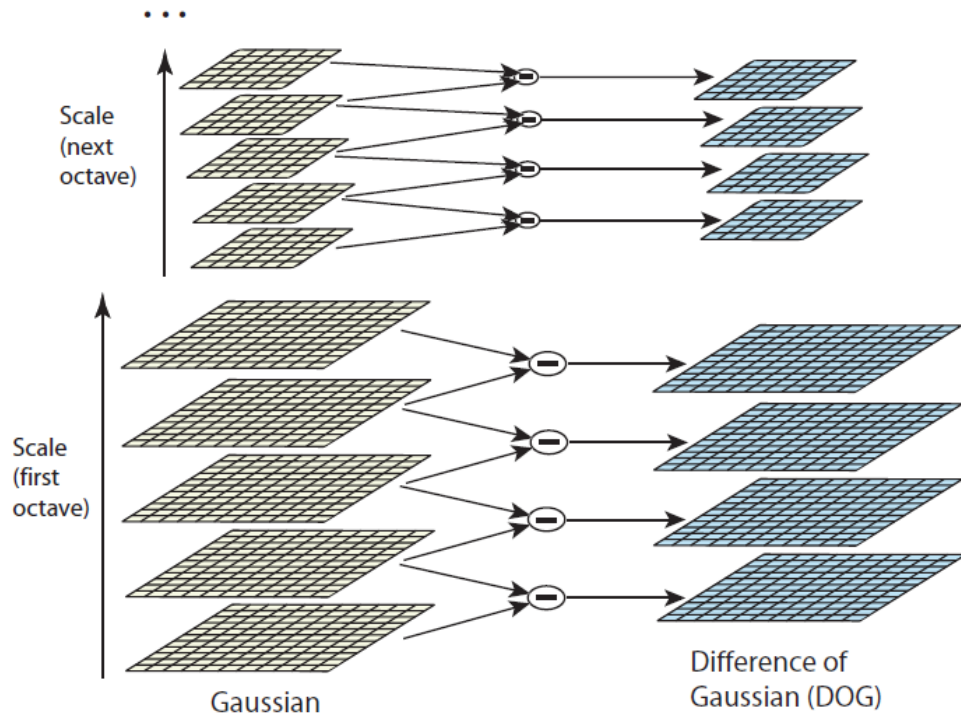
Figure 3. Within each octave, the image is convolved with Gaussians of increasing blurring strength, and then subtracted to create the difference of Gaussian images (Lowe, 2004).
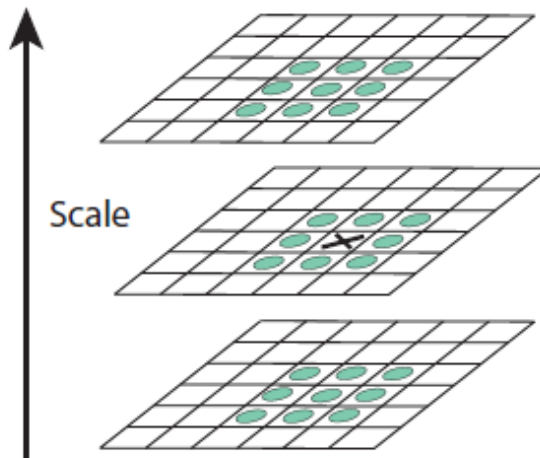
Figure 4. Extrema in the difference of Gaussian images are detected by comparing each pixel to the 26 neighboring pixels from the current and adjacent difference of Gaussian images (Lowe, 2004).

Once a local minimum or maximum is identified, a three-dimensional quadratic function is fit to the identified pixel to identify the interpolated actual location of the extremum as well as its scale. Besides location and scale, the final step to fully define a keypoint is to calculate its orientation. The SIFT algorithm calculates the direction of the largest gradient for pixels surrounding the extremum location. These orientations are weighted using a Gaussian window and then accumulated into an orientation histogram as shown in Figure 5 below.

Image gradients   Keypoint descriptor
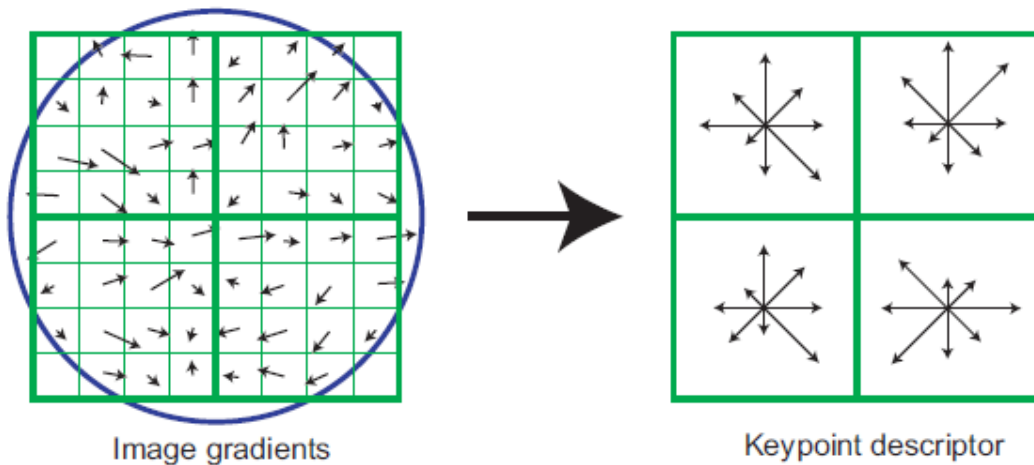
Figure 5. The neighborhood of gradient orientations and scales define the keypoint descriptor for each keypoint (Lowe, 2004).

The end result of this process is a uniquely constructed vector for each keypoint which contains the keypoint location, scale, and orientation. Once identified, a matching algorithm uses an approximation to the Euclidean distance to find keypoints that are similar between the two images. When the Euclidean distance between keypoint identifiers falls within the set threshold, a match is declared and the keypoint locations within each image are retained

3.3.3 Applying RANSAC to Cull Incorrect Matches

With the match points identified, the next step in the Walli registration process is the application of the RANSAC, or RANdom SAmple Consensus, algorithm to cull potential incorrect or poor matches. The RANSAC algorithm works by maximizing the number of "inlier" matches for which the fitting function holds true up to some threshold. The steps of the RANSAC algorithm are shown in Figure 6.

Objective

Robust fit of a model to a data set S which contains outliers

Algorithm

1. Randomly select a sample of s data points from S and instantiate the model from this subset.
2. Determine the set of data points S_i which are within a distance threshold t of the model. The set S_i is the consensus set of the sample and defines the inliers of S.
3. If the size of S_i (the number of inliers) is greater than some threshold T, re-estimate the model using all the points in S_i and terminate.
4. If the size of S_i is less than T, select a new subset and repeat the above.
5. After N trials the largest consensus set S_i is selected, and the model is re-estimated using all the points in the subset S_i.

Figure 6. The RANSAC algorithm applied to cull bad matches using the epipolar and homography constraints (Hartley and Zisserman, 2003).

As an example, RANSAC could be used to determine the best fit line for a set of two-dimensional data points. The fitting function in that case is y = mx + b, where x and y are the coordinates for each point, m is the slope, and b is the y-intercept. Thus, the set S would be the collection of the x and y coordinates in the dataset. The algorithm first randomly selects a pair of points and computes the perpendicular distance of all of the remaining points to the line generated by the two selected points. For each point, if the perpendicular distance falls above a threshold (t in step 2 of Figure 6), it is labeled as an

outlier. The algorithm stops once either a maximum number of iterations is reached, or the size of the inlier set, S_i, reaches a threshold T. T is generally determined from the measurement noise. Figure 7 below shows pictorially this example application of RANSAC to determine a linear fit.
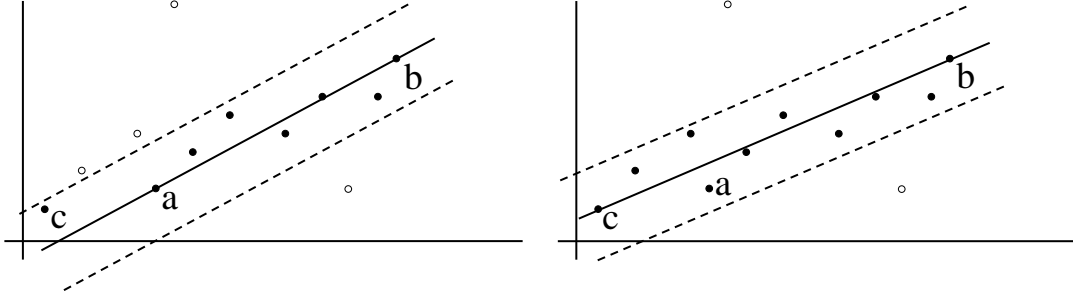
Figure 7. Application of RANSAC finding the best fit line for two-dimensional data (Hartley and Zisserman, 2003).

In the case of registering images in this research, the fitting functions used enforce the epipolar geometry constraint and homography constraint. For the former, the RANSAC algorithm is applied using Zisserman's 7-Point Algorithm for determining the fundamental matrix relating the images in epipolar geometry shown in Figure 8.
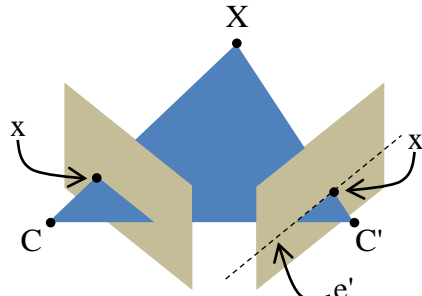
Figure 8. Epipolar geometry for two cameras (Hartley and Zisserman, 2003).

In Figure 8 there are two cameras shown with their camera centers at C and C' and with image planes indicated by the tan parallelograms. The matched three-dimensional point, capital X, is the real world point which both cameras have imaged.

This point projects to point x in the image plane of camera C and the line formed by x

and C, when projected onto the image plane of camera C', forms the epipolar line e'.

Thus, for any point x in the image plane of camera C, there is a corresponding epipolar

line e' in the image plane of camera C'. It is along this line which all x' must exist for

matching points. The fundamental matrix is defined as the matrix which maps the point x

to its epipolar line e' as seen in Equation (3.13):

$$Fx = e',$$
(3.13)

If both sides of Equation (3.13) are multiplied from the left by the transpose of x', the

following relationship results:

$$x'^T Fx = x'^T e' = 0,$$
(3.14)

since x' lies on e', the distance between the two ($x'^T e'$ is the dot product of x' and e') is

zero. Thus, the epipolar constraint enforced by RANSAC culls matches determined by

SIFT which do not conform to an epipolar geometry. RANSAC therefore determines the

best fit fundamental matrix F which maximizes the number of inlier points such that x'Fx

= 0 is true up to some threshold (i.e. x'Fx is less than some threshold) where x' and x are

the homogeneous coordinates of matched keypoints from the images under consideration.

To do this, RANSAC randomly selects seven matched points and calculates the

fundamental matrix relating them. It then tests the fundamental matrix against all of the

matched points to see how many matches fall within the threshold. The matches that are

under the threshold are called inliers and the rest are outliers. The algorithm iterates until

it finds a transformation matrix which includes enough inliers that the statistical

probability of the matches used to compute the fundamental matrix consisting entirely of

inliers reaches 99%. The output is the set of inliers for this 99% case (Hartley and

Zisserman, 2006). However, the application of RANSAC using the fundamental matrix

only ensures that matched points lie on the same epipolar line.

To further cull bad matches, RANSAC is reapplied using a two-dimensional

homography constraint in which all matched points result in the same transformation

matrix relating the homographic coordinates of the first image to the coordinates in the

corresponding image. The transformation matrix relating two two-dimensional images is

a three by three matrix and, since the imagery used in this research is all taken from the

nadir perspective and georectified, it is assumed that the transformation matrices can be

constrained to affine transformations and that the amount of three-dimensional ambiguity

between the images is small. Affine transformations can consist of a multiplicative linear

portion and an additive offset or translation portion which can be written as

$$[x \quad y] = [w \quad z] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + [b_1 \quad b_2], \qquad (3.15)$$

where w and z are the original image coordinates, x and y are the transformed image

coordinates, the a matrix is the multiplicative linear portion, and the b matrix is the

additive translation portion. Alternatively, and better for computational convenience, the

affine transformation can be written as a single matrix as

$$[x \quad y \quad 1] = [w \quad z \quad 1] \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ b_1 & b_2 & 1 \end{bmatrix}. \qquad (3.16)$$

Table 2 shows several types of affine transformations which can all be combined using

matrix multiplication.

Table 2. Common types of affine transformations (Gonzalez, Woods, Eddins, 2009).

Identity	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	
Scaling	$\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	
Rotation	$\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	
Horizontal Shear	$\begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	
Vertical Shear	$\begin{bmatrix} 1 & \beta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	
Horizontal Reflection	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	
Translation	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \delta_x & \delta_y & 1 \end{bmatrix}$	

Under the homography constraint, RANSAC randomly chooses sets of matches and determines the affine transformation matrix which relates the coordinate systems of the two images under consideration. As in the case of the fundamental matrix constraint, it iterates until a transformation matrix is found such that the probability of the set of points used to create the transformation matrix consisting entirely of inliers reaches 99%. Careful consideration must be used here as the two-dimensional homography requires a planar relationship between images and so can declare valid matches as outliers if there are significant terrain relief differences embedded in the image. This step eliminates any matches which satisfy the RANSAC fundamental matrix step of lying on the same

epipolar line, but do not satisfy the constraint that the transformation matrix relating the two coordinate systems should be the same regardless of the matches used in its determination.

3.3.4 RMSDE Minimization

The final component of the Walli registration method finds the set of matches resulting in a transformation matrix with an average RMSDE that falls below a user set threshold. The remaining matches after RANSAC are fed into MATLAB's cp2tform (control points to transformation matrix) function (MATLAB, 2010). The cp2tform function determines the best fit transformation matrix, again limiting to affine transformations, relating coordinate systems of the matched points. The RMSDE for each matched pair is then calculated using Equation 3.17:

$$RMSDE = \sqrt{(x - Tx')^2}, \qquad\qquad (3.17)$$

where x is the homogeneous coordinate of a match in one coordinate system, x' is the homogeneous coordinate of the match in the other coordinate system ,and T is the determined transformation matrix relating the two. If the average RMSDE for all of the pixels is higher than the user-specified threshold, the match with the highest RMSDE is thrown out and the algorithm restarts. This loop iterates until the average RMSDE falls below the user-specified threshold or until so many matches are thrown out that the transformation matrix can no longer be determined, in which case the threshold is set too low for the set of determined matches or not enough matches are determined in the first place. The flowchart in Figure 9 shows the Walli registration process pictorially.
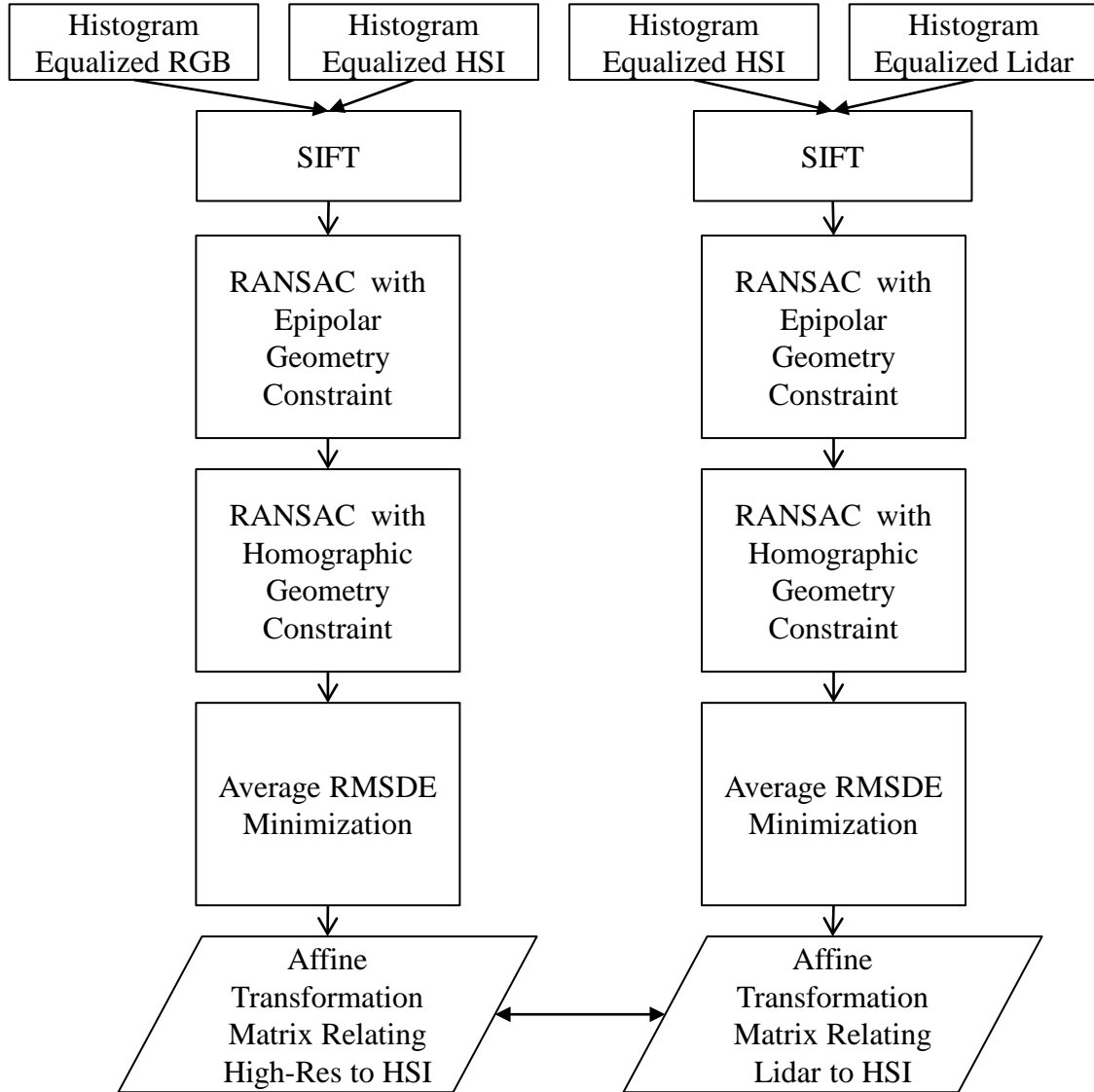
Figure 9. Flowchart of the Walli registration method used to register the three input modalities.

3.4   Chapter Summary

Automating the generation of three-dimensional scenes attributed with texture and

spectral information involves a number of various disciplines. This chapter attempts to

highlight the pieces most important to this work including the physics involved in the

DIRSIG software package and a thorough presentation of the Walli registration

techniques used to relate the input modalities. While previously used only to register

lidar and HR imagery, the Walli method is robust enough to include the HSI modality.

The method is therefore included in the LHD method as the first step to relate the three

input modalities.

IV. Automated Scene Extraction Methodology

4.1   Chapter Overview

This chapter presents the methods used to extract the geometry, texture, and spectral components of the synthetic scenes using the LHD method. Before the methodology is discussed, however, the two datasets used to develop and test the methods are presented. The first datset is a DIRSIG-generated HR RGB, lidar, and HSI dataset produced using a high-fidelity scene of the VanLare wastewater treatment plant near Rochester, New York. The second dataset is a real collect from the NEON prototype data collect over Melrose, Florida. These two datasets are used throughout the rest of the chapter to show example inputs and outputs for each step in the scene extraction process.

Next, the registration process is shown with results for each dataset followed by the methods for geometry and texture map extraction. Following the geometry and texture map extraction, the chapter presents three approaches to automated extraction of the material map. Approach 1 is a non-fused approach in which the HSI alone is used to generate the material map and spectral library. The second and third approaches attempt to improve the resolution and accuracy of the material map by fusing the HSI, lidar, and HR information. While it does not produce satisfactory results and is not fully automated, Approach 2 is included for its importance in developing Approach 3 and as a method which could be further developed in future research. The second fused approach, Approach 3, is a fully automated, fused method which produces HR material maps. The

final section presents the method used to extract the scene spectral library using the material maps resulting from Approach 1 and Approach 3.

4.2   Introduction to the Datasets

The DIRSIG-generated synthetic dataset consists of HR RGB imagery, HSI, and lidar data created by flying the respective simulated sensors for each of those modalities over a high-fidelity, three-dimensional scene of the VanLare wastewater treatment plant near Rochester, New York. Because the dataset is synthetic, perfect ground truth and knowledge of the sensor and its locations are available, and thus DIRSIG generated products using the autonomously recreated scene can be compared to the original DIRSIG products created using the original high-fidelity scene with identical sensor parameters, solar illumination, and atmospheric conditions. The second dataset consists of lidar and HSI from the NEON prototype collect over the Ordway Swisher Biological station near Melrose, Florida along with HR imagery over the same site provided by the Florida Department of Transportation (Krause and Kuester, 2011).

4.2.1   VanLare Wastewater Treatment Plant Synthetic Dataset

The VanLare dataset is a DIRSIG-generated dataset consisting of synthetic imagery and lidar data meant to be representative of the real NEON dataset introduced below with 0.3 meter GSD RGB imagery, an aerial HSI sensor with 3.0 meter GSD, and an aerial lidar sensor with approximately 1.0 meter posting. All imagery and lidar collected over the VanLare scene are captured from a nadir viewing geometry. The high-fidelity three-dimensional scene consists of the buildings, holding tanks, and surrounding forest and grass area of the VanLare wastewater treatment plant near Rochester, New

York. The scene was built by the DIRSIG development team using geometry drawn from an existing model of the site in Google Earth and enhanced with HR texture and material maps by Walli (Ientilucci and Brown, 2003; Walli, 2010). The buildings in the scene are attributed with HR material and texture maps for both horizontal and vertical surfaces ensuring good spectral accuracy for any viewing angle. Finally, spectral information is attributed by performing image segmentation on the texture maps and manually assigning material spectra from a spectral library populated with field spectra acquired throughout the scene (Walli, 2010). The modeled RGB, hyperspectral, and lidar sensors used to produce the synthetic images for input into the LHD method are modeled as idealized frame cameras without pointing or noise error. DIRSIG allows for the consideration of these parameters and future studies could investigate their effect of noise and pointing error on the LHD process. The synthetic images used for registration are shown in Figure 10.

4.2.2   NEON Prototype Collection over Ordway Swisher Biological Station

The National Ecological Observatory Network is a data collection project designed to assess climate change across the United States. The prototype HSI and lidar dataset collected by the NEON project in November of 2010 is available to the public. The HSI and lidar data from this collect along with HR imagery from the Florida Department of Transportation are used here to provide a test of the robustness of the LHD method as well as provide a qualitative and quantitative evaluation resulting autonomously synthetic scenes.

Figure 10. The input DIRSIG generated synthetic imagery. Top: HR imagery, Middle: Lidar elevation and return strength and Bottom: HSI captured over the modeled VanLare Wastewater Treatment Plant near Rochester, NY.

Both the lidar and HSI are from the Ordway Swisher Biological Station portion

of the NEON collect which includes parts of the town of Melrose, Florida. The lidar

collect has an approximately one meter posting and the HSI, collected by NASA's

AVIRIS hyperspectral imager and converted to reflectance using FLAASH, has a 3.4

meter GSD. AVIRIS has 224 bands across the 400 to 2500 nm spectral range with each

band spanning approximately 10 nm. The HR imagery is provided by the Florida

Department of Transportation website and has an approximate GSD of 0.3 meters. The

original images used for registration are shown in Figure 11 below.
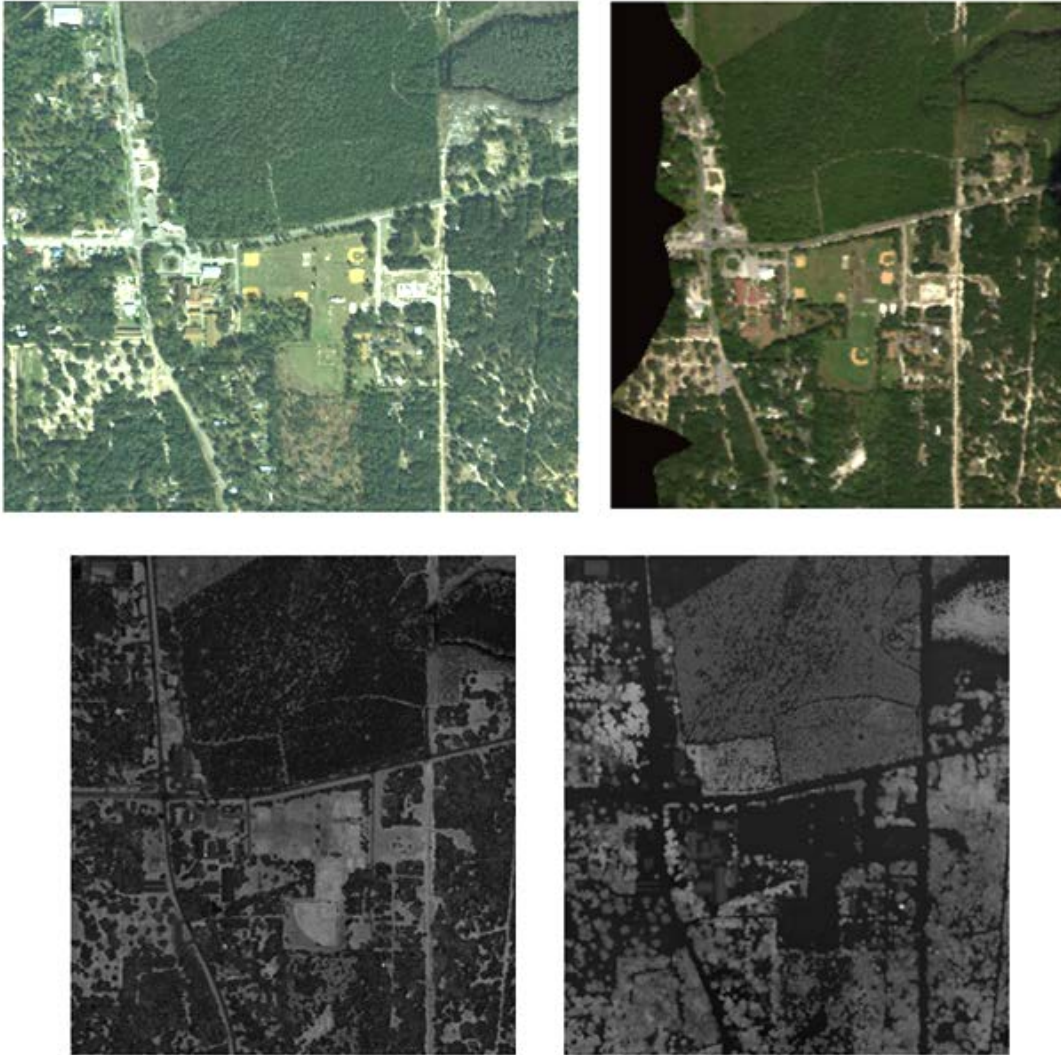
Figure 11. The NEON input dataset captured near Melrose, Fl. Clockwise from top left: HR RGB imagery, AVIRIS HSI, lidar return strength, and lidar elevation.

4.3   Image Registration

The input modalities are prepared for registration by following the image

preparation flowchart shown in Figure 2 of Chapter III. For the lidar, only the first

returns of the lidar are used and the data consist of the X, Y, Z coordinates in UTM

format along with the infrared (IR) return strength. The IR returns are rasterized using

MATLAB's TriScatteredInterp function resulting in an intensity image at the

illumination laser wavelength of 1064 nm. The higher resolution inputs (RGB and

rasterized lidar) are then resampled to the HSI Ground Sample Distance (GSD) using a

bilinear interpolation. The RGB image is then converted to grayscale using MATLAB's

rgb2gray function (MATLAB, 2010). Representative spectral bands re then selected

from the HSI (~450 nm, ~550 nm, and ~650 nm for the RGB image and ~1060 nm for

the rasterized lidar return strength). Next, the red, green, and blue HSI bands are

converted to grayscale using the same process used for the HR imagery. Finally all

inputs are histogram equalized to a flat histogram using MATLAB's histeq function

(MATLAB, 2010). The resulting grayscale images prepped for registration are shown in

the top row of Figure 12 below.

Following the Walli registration flowchart in Figure 9 of Chapter III, the second

row of Figure 12 shows the SIFT matches for both the high-res to HSI and lidar-to-HSI

registrations and the bottom row shows the surviving matches (in green) after culling by

RANSAC using the fundamental matrix and homographic constraints followed by the

average RMSDE minimization to under 0.25 pixels average RMSDE (at the HSI GSD).

The resulting transformation matrices for each registration are shown below in Table 3

along with the number of remaining matches after culling.

Following the same image preparation procedures for the Melrose collect, the

FLDOT RGB imagery is registered to similar red, green, and blue HSI bands and the

rasterized lidar return strength is registered to the HSI band nearest the 1064 micron lidar

illuminating laser.
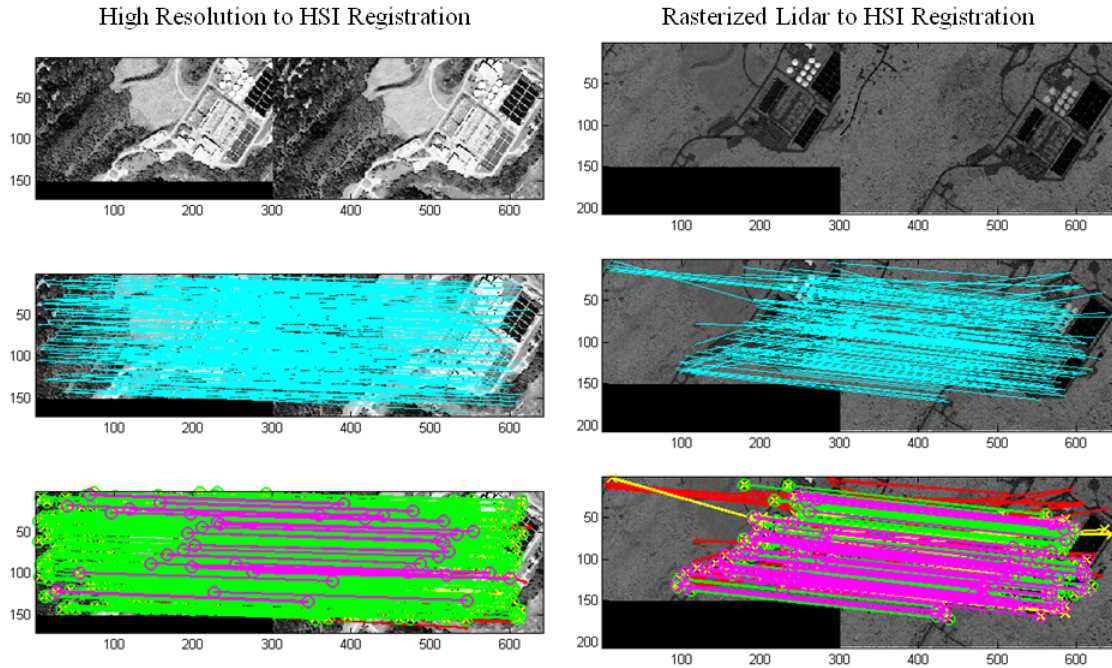
Figure 12. Left: Each step of the HR to HSI registration process showing the original images side by side (top), match points detected by SIFT in light blue (middle), and remaining good matches in bright green after culling (bottom), Right: The same steps for the lidar-to-HSI registration.

Table 3. HR-to-HSI and lidar-to-HSI registration results

High res to HSI Transformation Matrix			Lidar-to-HSI Transformation Matrix		
0.9867	-0.00003	0	1.0084	0.0003	0
-0.00003	0.9867	0	-0.0003	1.0084	0
-18.4149	-9.2216	1	-28.0693	-35.5933	1
RMSDE Average: 0.2483 pixels			RMSDE Average: 0.2453 pixels		
Matches Remaining: 306			Matches Remaining: 49		

The two registrations are shown in Figure 13 below with the RGB to HSI registration on

the left and lidar-to-HSI registration on the right. The resulting registration matrices and

match information are shown in Table 4.

49

High Resolution to HSI Registration   Rasterized Lidar to HSI Registration
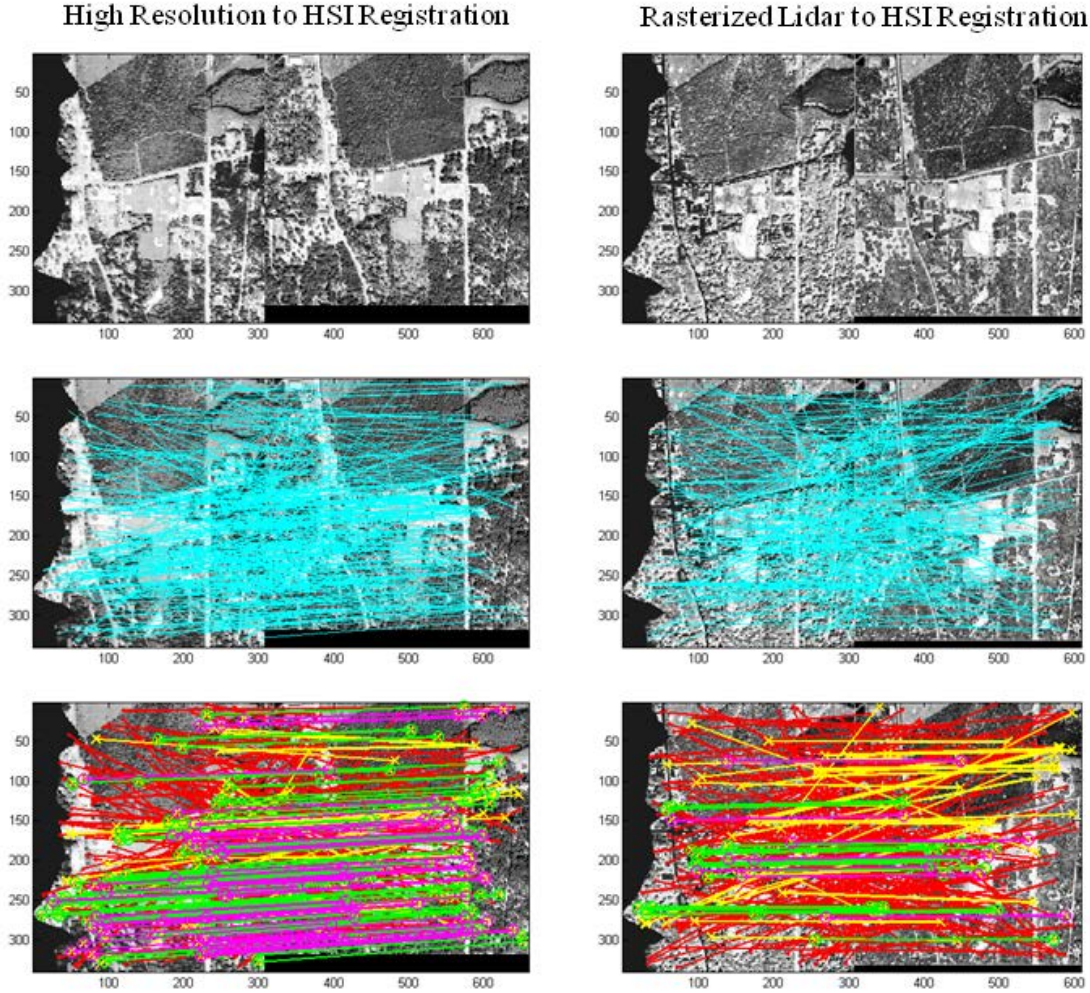
Figure 13. Left: Each step of the HR-to-HSI registration process showing the original images side by side, match points detected by SIFT in light blue, and remaining good matches after culling in green, Right: The same steps for the lidar-to-HSI registration.

Table 4. HR-to-HSI and lidar-to-HSI Registration Results

High-res to HSI Transformation Matrix			Lidar-to-HSI Transformation Matrix		
0.9103	0.0087	0	0.9671	-0.0091	0
-0.0113	1.0500	0	-0.0029	0.9658	0
-10.6637	7.9622	1	0.8411	12.0426	1
RMSDE Average: 0.4996 pixels			RMSDE Average: 0.4800 pixels		
Matches Remaining: 58			Matches Remaining: 20		

Transformations from one image coordinate system to another are all applied using

bilinear interpolation unless transforming from the HSI coordinate system . For

50

transformations from the HSI coordinate system, nearest neighbor interpolation is used to prevent the spectral component from being polluted with spatial variation.

4.4 Extraction of Scene Geometry

As discussed in the DIRSIG Scene Components section of Chapter III, DIRSIG scenes consist of facetized three-dimensional geometries. For this research, the geometries are input into DIRSIG using the common .obj file format also known as the wavefront format. This format consists of a list of the positions of all of the vertices followed by a list containing the vertex indices which make up each facet in the scene. Initially, these geometry files were generated directly from the input lidar point clouds using a Delaunay triangulation function in MATLAB to generate the facet list. However, it became apparent that this simple tinning technique, while sufficient for ground and solid objects within the scene, is not sufficient for handling objects like tree canopies which should allow some light to pass through. Thus, it is necessary to identify and separate ground and building lidar returns from tree canopy lidar returns to allow a different treatment of the tree canopies.

The tree returns are separated from ground returns by first rasterizing the lidar point cloud then performing a morphological opening using MATLAB's imopen function with a 20 pixel radius disk as the structuring element. The morphological opening of a grayscale image, like the rasterized lidar elevation map, eliminates peaks which are too narrow for the structuring element to fit into. Figure 14 shows pictorially the morphological opening of a grayscale "image" in one dimension.

51
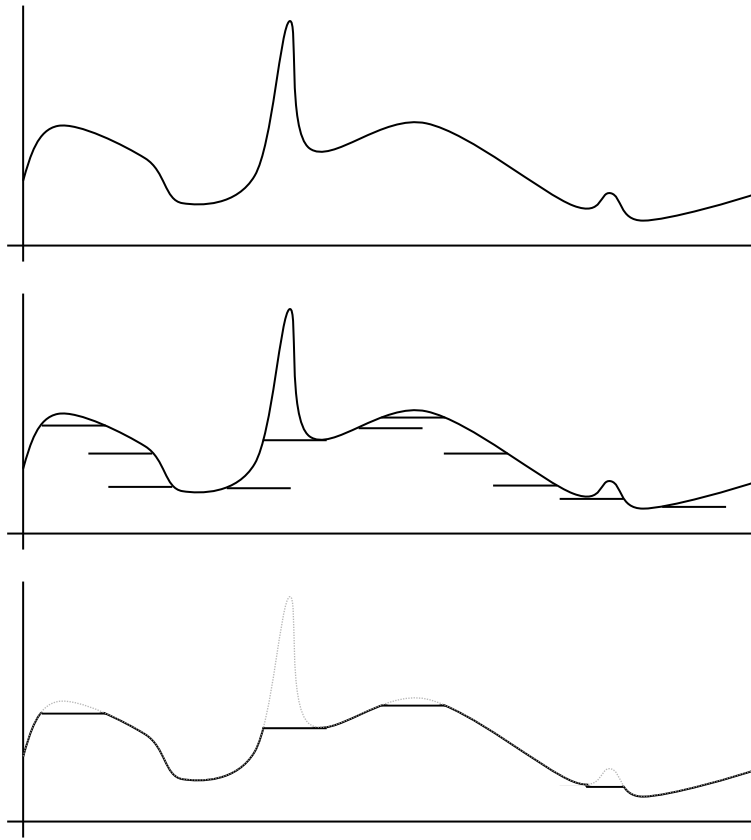
Figure 14. Demonstration of a grayscale morphological opening in one dimension.

The result is a ground terrain map which is then subtracted from the original rasterized elevation map to create an above ground elevation map including any trees and structures in the scene which are smaller than the structuring element. Trees are separated from buildings by taking advantage of the higher local standard deviation over vegetation when compared to man-made structures. An above ground standard deviation map is created using MATLAB's stdfilt function and then thresholded to generate a tree mask (MATLAB, 2010). The ground and building map is then generated by replacing tree pixels in the original elevation map with pixels from the ground terrain map. The ground and building map is then facetized using the "Delaunay" function and the tree map is used to generate half-meter square tiles to represent trees. The separated geometries are

then written out into the standard .obj format which DIRSIG can use directly. Figure 15

shows the geometry extraction process in flowchart form.
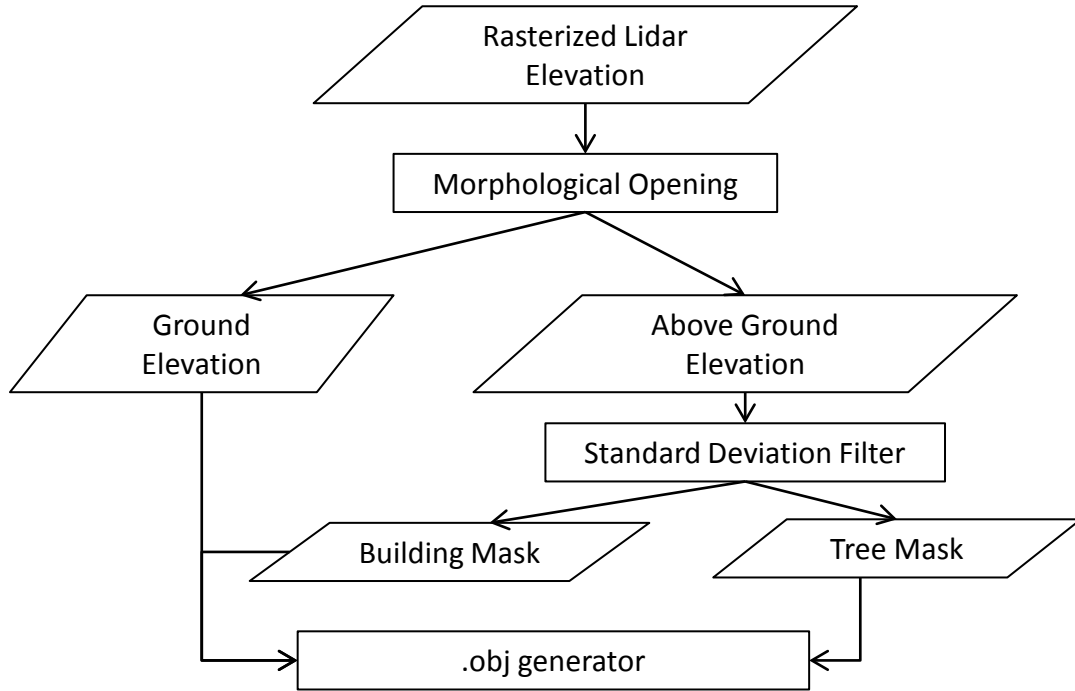
Figure 15. The geometry extraction process developed for the LHD method.

Using the process above, the lidar point cloud is rasterized and tree pixels are

extracted using a morphological opening with a disc-shaped structuring element of 20

pixel radius and then isolated from building pixels by thresholding a standard deviation

image at 0.5 meters. Figure 16 shows the full rasterized elevation map of the VanLare

dataset (top left), followed by the ground terrain map resulting from the morphological

opening (top right), the above ground terrain map from the subtraction of the ground

terrain map from the full elevation map (bottom left), and the tree mask resulting from

the standard deviation threshold of the above ground terrain map (bottom right).
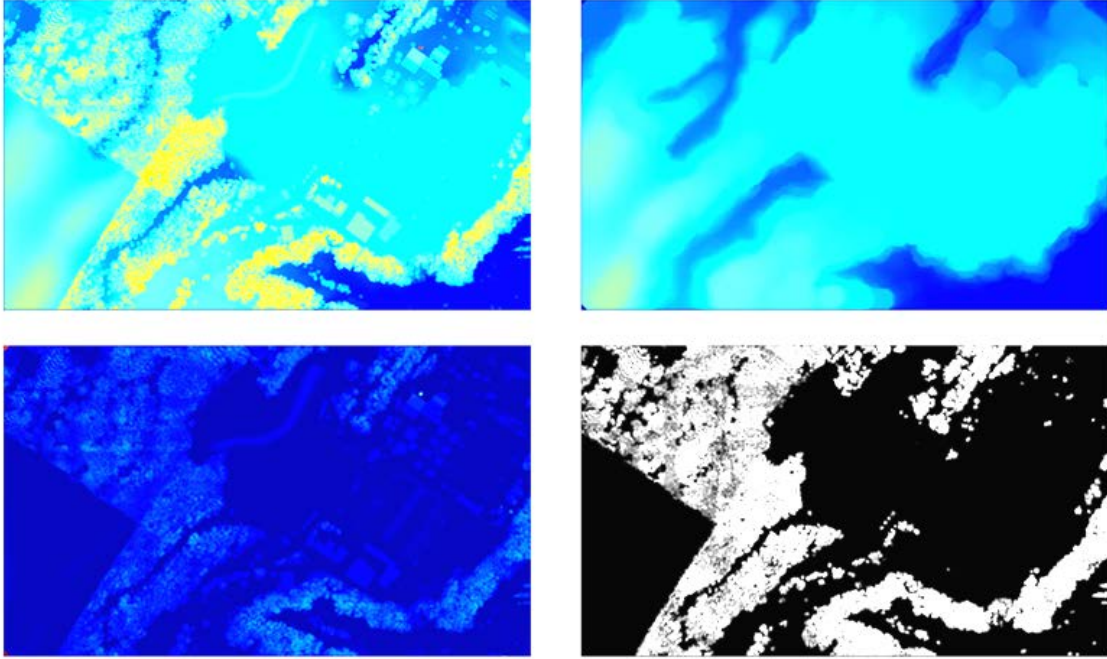
Figure 16. Top left: The full rasterized lidar elevation, Top right: The ground terrain map, Bottom left: The above ground terrain map, Bottom right: The tree mask used to extract tree pixels from the original full elevation map.

The x, y, and z coordinates are then output into separate point clouds: One for the ground and buildings and one including only the trees. The ground and building point cloud is facetized and written out into standard .obj format. At each tree return, a half-meter square facet is added to the .obj file. While a half-meter facet size is used here, appropriate adjustment would be required for higher or lower resolution input lidar data. The resulting geometry, rendered using Meshlab (Visual Computing Lab, 2014), is shown in Figure 17. The resulting geometry for the Melrose scene using the same extraction approach is shown below in Figure 18 as a Meshlab rendering (Meshlab, 2014).

Figure 17. The resulting geometry produced from the synthetic "original" images written in obj format and rendered using Meshlab.
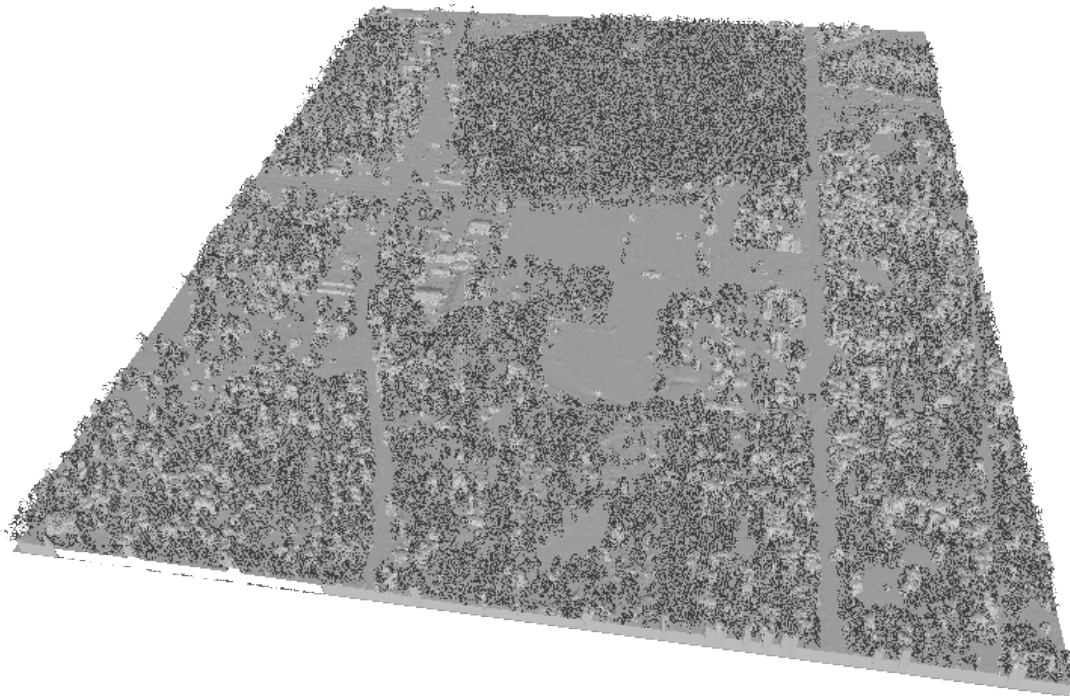
Figure 18. Meshlab-rendered geometry extracted from the NEON lidar collected of Melrose, Florida with tree returns separated from ground returns and modeled as half-meter tiles.

4.5   Extraction of Scene Texture Maps

The most easily generated of the three scene components is the texture component. The texture component is drawn directly from the HR RGB image bands which are transformed to lidar space using the registration results and bilinear interpolation. The transformed images are then draped over the scene geometry. DIRSIG allows users to set up multiple texture maps to cover different spectral regions. Thus, three texture maps are generated here using the red, green, and blue bands of the HR imagery to create red, green, and blue texture maps covering the 0.6 μm to 0.7 μm, 0.5 μm to 0.6 μm, and 0.4 μm to 0.5 μm regions respectively. DIRSIG also allows users to specify material specific means and standard deviations for each texture map. These numbers are extracted by transforming the material map (generated using one of the approaches presented below) to the HR space using the registration transformation matrix and nearest-neighbor interpolation to identify material regions of interest for which intensity means and standard deviations for each material class are calculated. The final step in preparing the texture maps is to calculate the insertion point into DIRSIG. This is the upper left point of the image transformed to the lidar coordinate system and can be found by multiplying the (1,1) coordinate by the series of affine transformations used in the image preparation and found in the registration process. Appendix 1 provides a detailed explanation of how to calculate the insertion points for the texture and material maps. The resulting texture maps for the VanLare dataset are shown in Figure 19 and for the Melrose dataset in Figure 20.

Figure 19. Left to right: The red texture map, green texture map, and blue texture map transformed to the lidar coordinates to be draped over the scene geometry in DIRSIG.

Figure 20. Left to right: The red, green, and blue texture maps generated by transforming the FLDOT RGB imagery to lidar space using the registration results.

4.6   HSI-only Material Map Extraction: Approach 1

This section details the first of three methods attempted in this research to fully automate the generation of a three-dimensional scene attributed with texture and spectral information (Givens, et al., 2012). Unlike Approaches 2 and 3 which are presented in the next two sections and use a fused approach, Approach 1 uses the hyperspectral input alone to generate a material map and associated material library. It uses a common unsupervised classification routine called the Stochastic Expectation Maximization (SEM) method which employs a randomized approach to classification by Mahalanobis distance thereby allowing some class overlap. A more detailed description of the mathematics behind the SEM algorithm is beyond the scope of this dissertation, but it has been shown to be highly effective in hyperspectral data processing (Eismann, 2012).

Besides being selected for its common use in the field, SEM is fully unsupervised meaning that it requires no user input to perform scene classification and it therefore fulfills the goal of this research to fully automate the generation of DIRSIG scenes. In general, the strategy for accomplishing spectral image classification consists of feature extraction, spectral classification, and class labeling with the end result being the material map as shown below in Figure 21.

Sensor Data → [Feature Extraction] → [Spectral Classification] → [Class Labeling] → Classification Map

Figure 21. Process of extracting a material map from sensor data (Eismann, 2012).
However, since a material can be designated simply by a number or some semantic name like grass or tree for DIRSIG, and the material spectra are pulled directly from the HSI, the final step of class labeling, at least semantically, is not necessary. If material information like bulk properties, BRDF, or other information not available directly from the HSI were to be included, the class labeling step would be required in the form of matching spectra from the HSI to some spectral library where the extra information is available.

The SEM algorithm runs on the hyperspectral image cube alone. SEM is a quadratic clustering algorithm, meaning it assumes pixels (spectra) within hyperspectral imagery fall within clusters in some N-dimensional space using some distance metric and that the decision line (or N-dimensional surface) separating the classes can be described by a quadratic equation as illustrated in Figure 22.
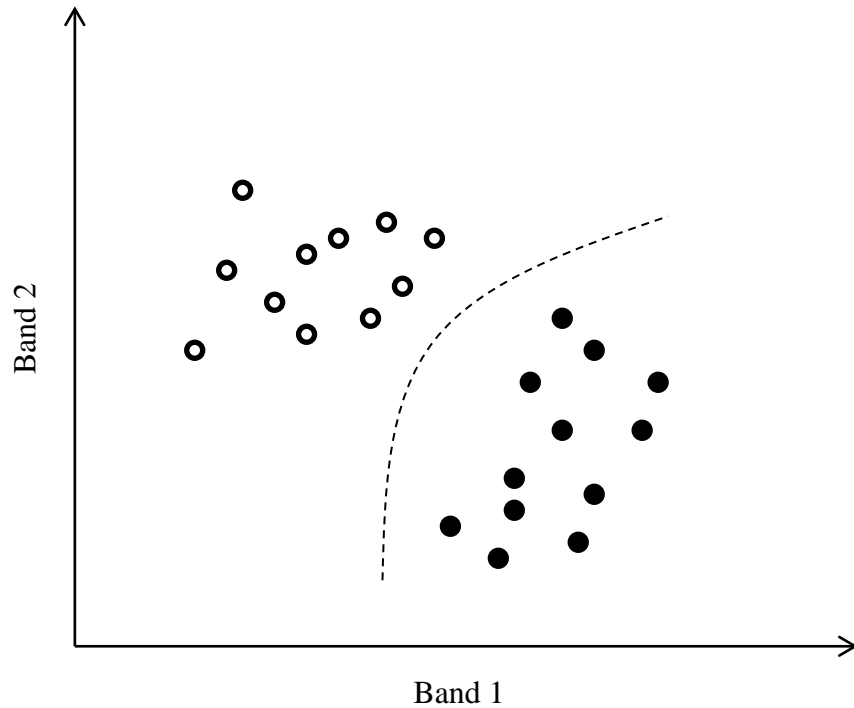
Figure 22. Illustration of quadratic classifier (Eismann, 2012).

Before running the SEM algorithm, Approach 1 uses principal components

analysis (PCA) to reduce the dimensionality of the hyperspectral cube both to improve

performance of the SEM algorithm by choosing axes which maximize the variance

between bands and thus the distance between clusters, and to reduce the number of bands

used in the distance calculation and thus reduce processing requirements. The PCA is

useful when dealing with HSI due to the large correlation between spectral bands (i.e.

changes in radiance or reflectance over the small spacing of HSI spectral bands is

generally small). The PCA computes the eigenvalues and eigenvectors of the correlation

matrix for the HSI cube and then projects the HSI cube onto those eigenvectors in order

of decreasing eigenvalue (and thus decreasing variance captured per band). The result is

a new principal component cube in which bands are both uncorrelated and sorted in order

of decreasing variance. Additionally, by adding the eigenvalues of some subset of the

59

principal component cube and dividing by the sum of all of the eigenvalues, the percent

of variance captured in the reduced set is known. Again ,due to the high correlation

between spectral bands of HSI, a generally much smaller number of the PCA bands can

be used for operations like classification while retaining nearly all of the variance of the

original HSI cube. This process is known as dimensionality reduction (Eismann, 2012).

The only input to the SEM algorithm, besides the dimension-reduced HSI, is the

initial number of classes in which to assign the spectra. This number is generated from

the linear mixing model rule of thumb that the number of classes in the scene is equal to

the number of significant principal component bands plus one (Eismann, 2012). While

this assumption is based on the linear mixing model, which is not used in the Approach 1

classification of the scene, this common dimensionality reduction technique at least gives

a reasonable estimate for the number of materials in the scene. Determining the number

of significant bands could be approached using statistical methods to estimate noise in the

data, but is accomplished here by simply thresholding the amount of variance captured to

99.95%. Future research could potentially eliminate the requirement for this user-

specified thresholding of the variance by determining the threshold from an estimation of

the noise in the data (Chang and Du, 2004; Eismann, 2012).

For the VanLare scene, the 99.95% variance captured threshold is met by

retaining the first seven principal component bands, representing a maximum of eight

classes. The plot in Figure 23 shows the cumulative percent variance captured for each

of the first ten principal component bands in the VanLare HSI.
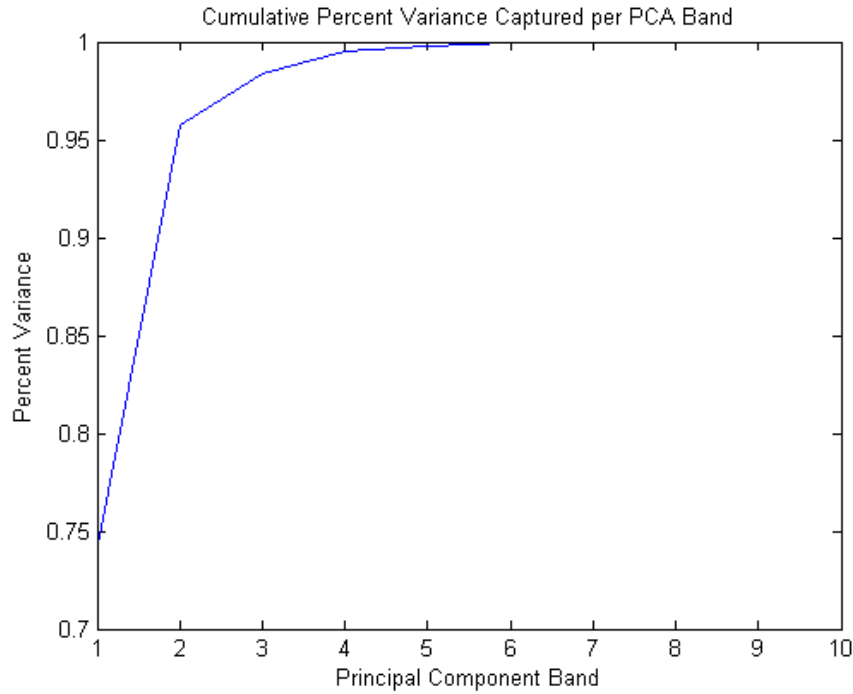
Figure 23. Cumulative percent variance captured for each of the first ten principal component bands of the VanLare HSI.

The SEM algorithm is then applied to the reduced dimensionality PCA cube. With the VanLare HSI reduced to seven principle component bands and the need for eight material classes identified, the inputs for the SEM algorithm are ready. Once the algorithm runs, the output is a material map at the GSD of the HSI shown in Figure 24. This material map is then input into DIRSIG after performing the image transformation resulting from the lidar-to-HSI registration. The flowchart showing each of the steps in Approach 1 is displayed in Figure 25.

Applying the same Approach 1 to the Melrose scene, the first step is the dimensionality reduction using principal component analysis to determine the initial number of classes for the SEM algorithm. For the AVIRIS HSI collected over the Melrose site, 14 principal component bands were needed to capture the 99.95% variance threshold. Thus, 15 classes were used to initiate the SEM algorithm.
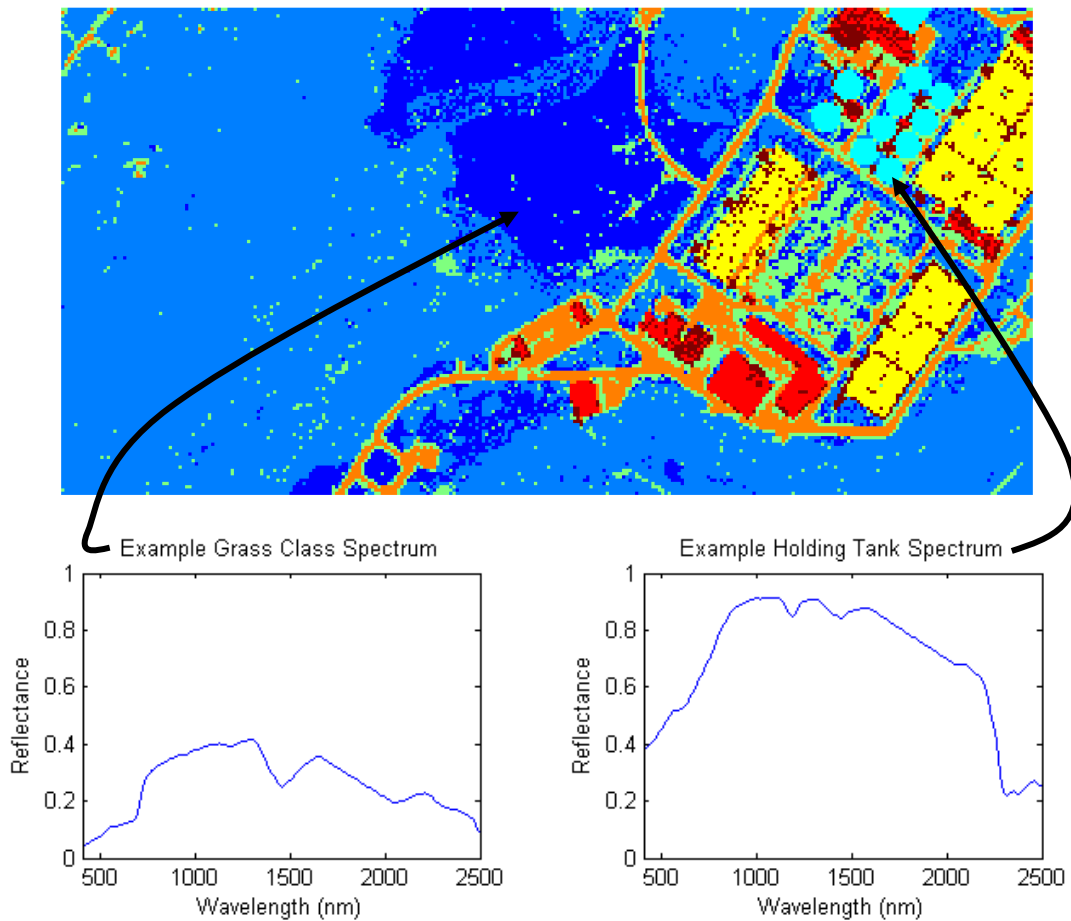
Figure 24. SEM material map and example mean spectrum for one of the classes.
The resulting material map is shown below in Figure 26. Note that the darkest blue color

on the left of the image is masked from the SEM algorithm and not assigned to a class

since the HSI is not available there.

While the non-fused Approach 1described above successfully automates the

generation of a material map for a synthetic scene, it fails to take advantage of the

information of the HR imagery and lidar data to potentially improve the classification

resolution and accuracy.
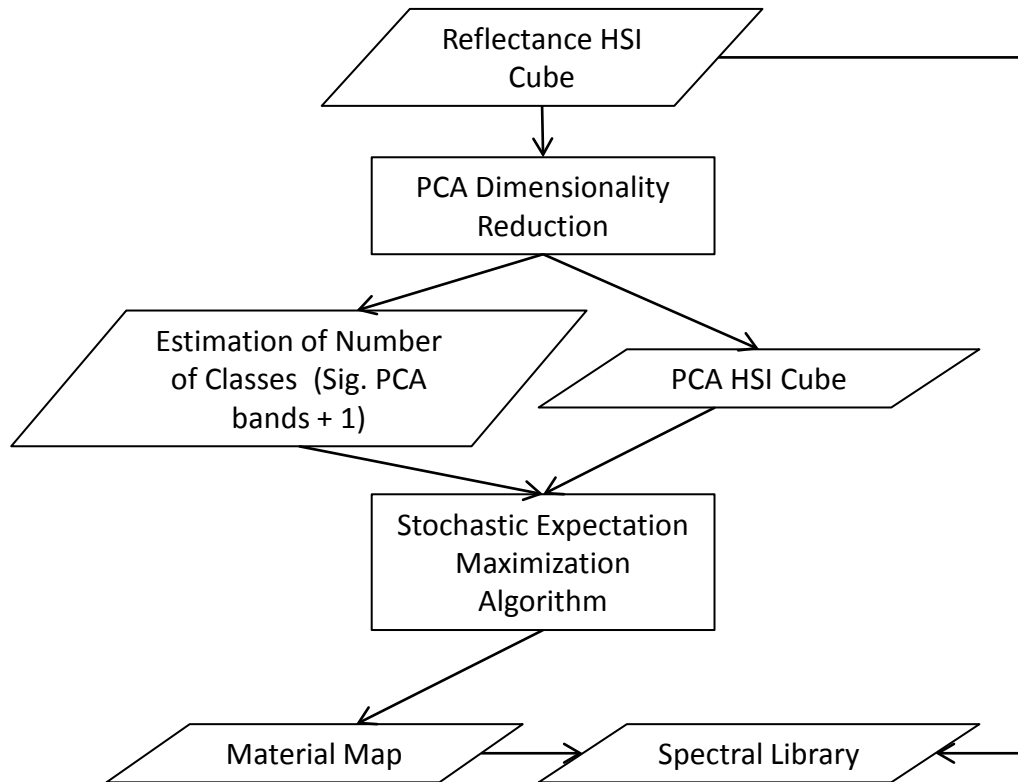
Figure 25. Flowchart of Approach 1.
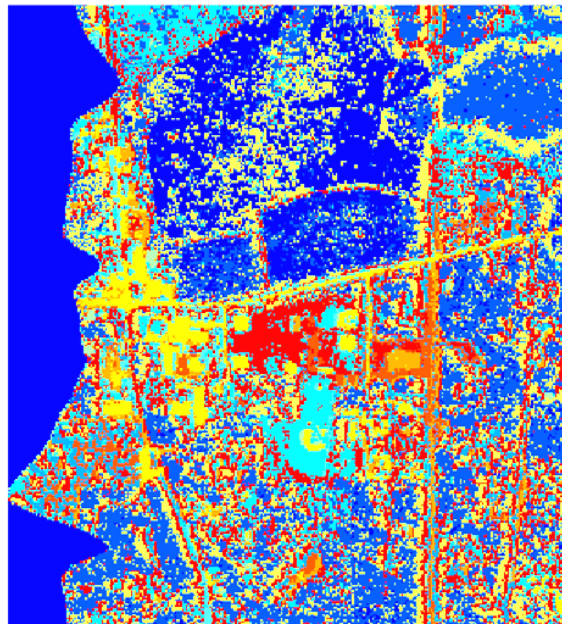
Figure 26. The material map resulting from the application of SEM on the input AVIRIS HSI using 15 initial classes.

63

With the material map at a much lower resolution than the scene texture maps and geometry, output synthetic imagery created using the low-resolution material map tends to show pixelation if the modeled sensor GSD is smaller than that of the input HSI. This pixelation is particularly evident along the edges of sharp material transitions like the grass and dirt edges in the playing fields of the Melrose scene. Since the other two higher resolution modalities are available and registered to the HSI in the LHD method, possibilities exist for the sharpening of the HSI material map.

4.7   Fused, Least Squares Unmixing Material Map Extraction: Approach 2

The second approach to scene classification performs an unmixing of the hyperspectral cube and then relates material percentages within each HSI pixel to spatial information found in the HR imagery and rasterized lidar elevation and return strength (Givens, et al., 2013). This method applies the linear mixing model in which each pixel in the scene is assumed to be a linear combination of all of the pure materials (endmembers) in the scene. Ideally, the abundance coefficients for each material should fall between 0 and 1, representing the percentage of each endmember material in the pixel, with the sum of the coefficients adding to 1. This model is appealing due to its intuitive description of the physics, but unfortunately is difficult to apply using real data because of the typically unavailable perfect endmember spectra in addition to sensor noise and other non-linear sources of error like scattering and adjacency effects.

While the goal of this research is to fully automate the process to build a DIRSIG scene, various automated approaches to selecting the endmembers for Approach 2 failed for both the VanLare and Melrose datasets. Two common algorithms were tested to attempt this feat: PPI and N-FINDR. However, neither produced endmember sets which

64

appeared logical for the scene or resulted in satisfactory unmixing. The PPI and N-FINDR algorithms were likely not able to determine satisfactory endmembers with the Melrose scene either because non-linear effects like scattering, sensor noise, and adjacency effects exist in the scene and thus the Linear Mixing Model is a poor model for the scene, or pure endmember spectra do not exist in the scene. Thus, the endmembers presented here are chosen by hand with the justification that the automation of the endmember selection process would be further investigated if the rest of the unmixing and fusion process produced good results. Endmembers are selected from the radiance HSI cube with the intent of capturing the major materials in the scene such as grass, tree, asphalt, and roofing material. Manually locating pure pixels is made easier with the help of the registered HR imagery and lidar elevation information. After endmembers are selected, per pixel abundances are calculated using the constrained linear least squares solver in MATLAB, lsqlin (MATLAB, 2010). The constraints limit abundances (the linear coefficients) to be positive and sum to one for every pixel in the scene.

Next, HSI pixels consisting of an endmember abundance greater than or equal to 90% are assumed to be pure, or near pure, pixels. These pixels are then used as regions of interest (ROI) to pull out elevation and color information from the lidar and HR imagery, respectively. Over each endmember ROI, elevation, return strength (from the registered, rasterized lidar images), and red, green, and blue (from the registered HR image) values are averaged to create a five component material description vector. Each HR pixel of the fused HR and rasterized lidar return strength and elevation cube is then assumed to be a pure pixel and classified as one of the endmembers using Spectral Angle Mapper (SAM). To maintain the spectral content of the original HSI, the pixel

65

assignments at the high resolution are constrained to match the abundances found during

the least squares fitting. However, due to fitting error with these datasets, this constraint

causes otherwise pure areas of pixels to contain some noisy, and likely incorrectly

assigned, pixels. Thus, the resulting material map is smoothed using the morphology

techniques of opening and closing, which are often used together for image smoothing

and noise removal. Figure 27 shows a flowchart of the main steps used in Approach 2.
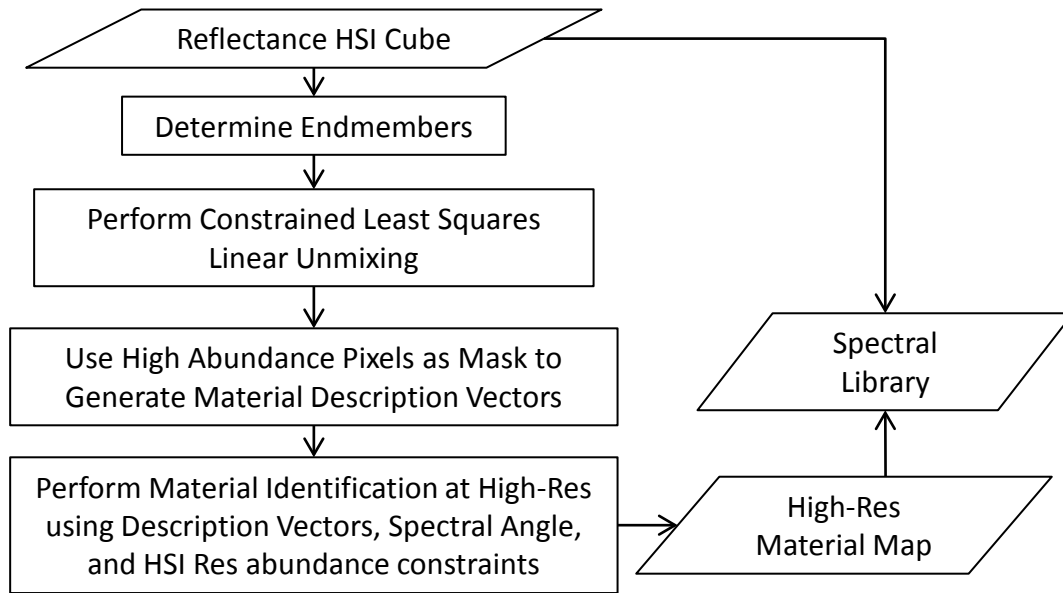
Figure 27. Flowchart showing the main steps of Approach 2.

To limit processing time while during the investigation of Approach 2, the input

modalities are cut down to the more manageable region of interest shown below in Figure

28. The region is selected for its spatial variability in materials and its range in elevation

(the trees and buildings) to show the value of the added information from the fused HR

imagery and lidar data.
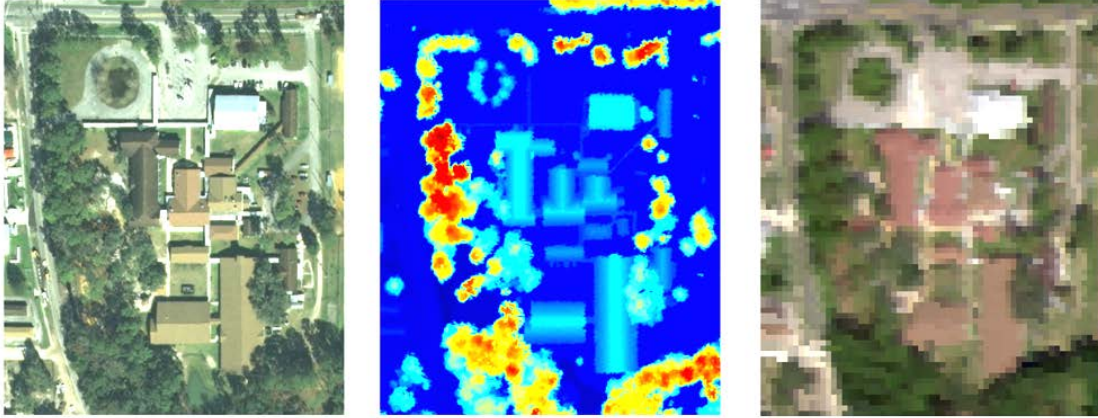
Figure 28. Left to right: Florida DOT HR imagery, rasterization of the NEON lidar elevation, and color composite of the NEON HSI showing the reduced ROI used to investigate the linear unmixing fusion approach.

As previously mentioned, this Approach is not yet fully automated and thus endmembers are selected by hand using the information in the registered modalities. For this example subset of the Melrose scene, eight endmembers are hand-selected: grass, tree, asphalt, parking lot, red roof, brown roof, and white roof with and without glare. Figure 29 below shows the radiance plots of the eight endmembers.
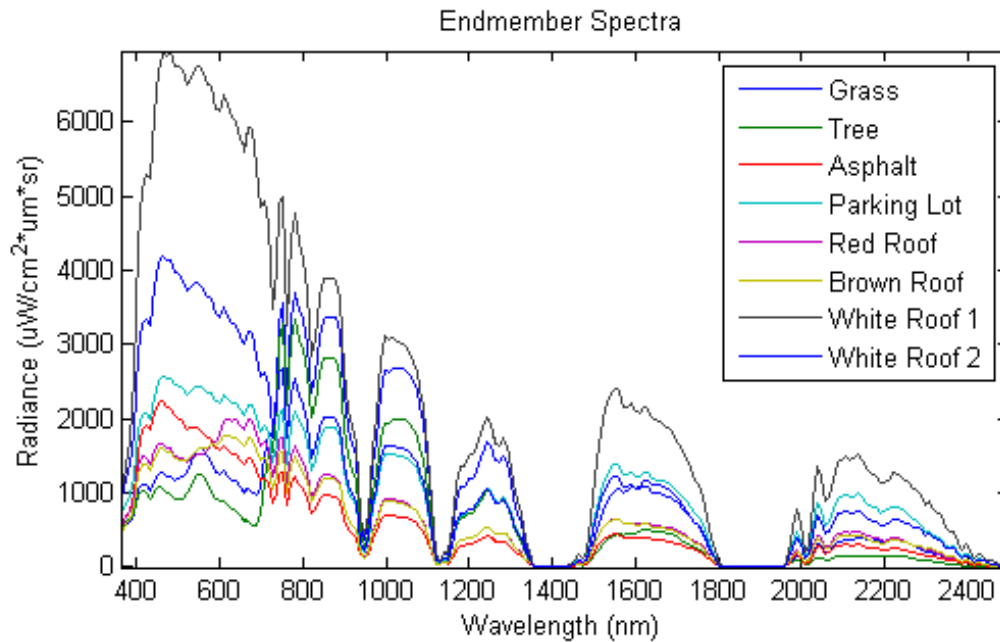
Figure 29. The Radiance spectra of the eight hand-selected endmembers used to perform linear unmixing of the HSI.

67

The eight endmembers are then input into MATLAB's linear least squares solver, linlsq with the constraints that all abundances are positive and sum to one. The resulting abundance maps are shown below in Figure 30. The second row of the figure shows where abundances are higher than 90%. These are considered pure pixel regions and used to define the ROIs of each of the endmembers which are then used to pull endmember attributes out of the HR and rasterized lidar imagery.
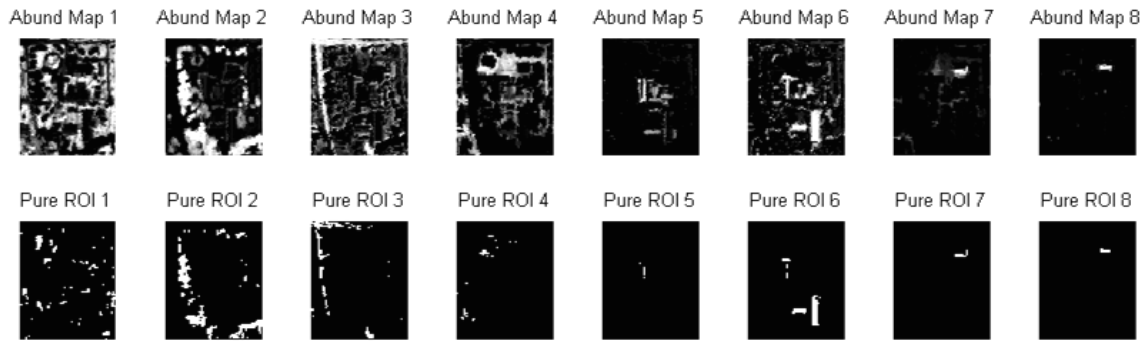
Figure 30. The top row shows abundance maps for each of the eight endmembers where numbers 1 through 8 correspond to endmembers grass, tree, asphalt, parking lot, red roof, brown roof, white roof with glare, and white roof without glare, respectively. The bottom row shows the pure pixel ROIs where each of the abundance maps contain pixels with abundances greater than 90%.

New material description vectors are defined by averaging over the HR and rasterized lidar values for the pure pixels of each class. The result is a five component description vector for each material class consisting of a red, green, and blue intensity value from the HR imagery and an elevation and return strength value from the rasterized lidar imagery averaged over the pure ROIs for each material. Table 5 shows the five components for each of the endmember-based materials.

With new material vectors calculated, each five-component pixel vector from the fused HR and lidar elevation and return strength cube are compared with and assigned to the endmember it most resembles.

68

Table 5. Material attributes calculated by averaging over HR red, green, and blue values and rasterized lidar elevation and return strength values.

	Red (intensity)	Green (intensity)	Blue (intensity)	Elevation (m)	Return Strength (intensity)
Grass	92.1	109.6	81.8	43.1	55.1
Tree	59.7	78.5	63.3	87.7	19.5
Asphalt	88.5	105.5	91.5	40.4	24.3
Parking Lot	128.9	145.5	125.0	40.7	35.6
Red Roof	82.5	88.5	74.0	56.5	17.1
Brown Roof	106.8	111.1	83.5	58.8	17.1
White Roof (with glare)	232.3	243.9	241.9	61.2	49.0
White Roof (without glare)	145.0	177.2	179.1	62.0	44.9

The method used for this step is Spectral Angle Mapper. For every pixel in the fused

HR, lidar elevation, and lidar return strength cube, the angle between its five component

vector and the five component vector of each of the eight endmembers is calculated.

Where the smallest angle occurs, the pixel is assigned as that endmember in a final

material map with the constraint that the final abundances of the HR material map must

average over the area of the corresponding HSI pixel to equal the abundances calculated

by the linear least squares unmixing used on the HSI alone. This final constraint ensures

that, on average, the subpixel results cannot change with respect to the spectral content of

the original HSI. The constraint is met by assigning pixels in order of least abundant to

most abundant endmember in the HSI pixel. For example, if a single HSI pixel mixture

is found to be 5% grass and 95% tree and that HSI pixel covers a ten-by-ten pixel area at

the high resolution, the five pixels with the smallest difference in spectral angle from the

grass class are assigned as grass and the remaining 95 pixels are assigned as tree. The

resulting material map using this method is shown below in Figure 31.
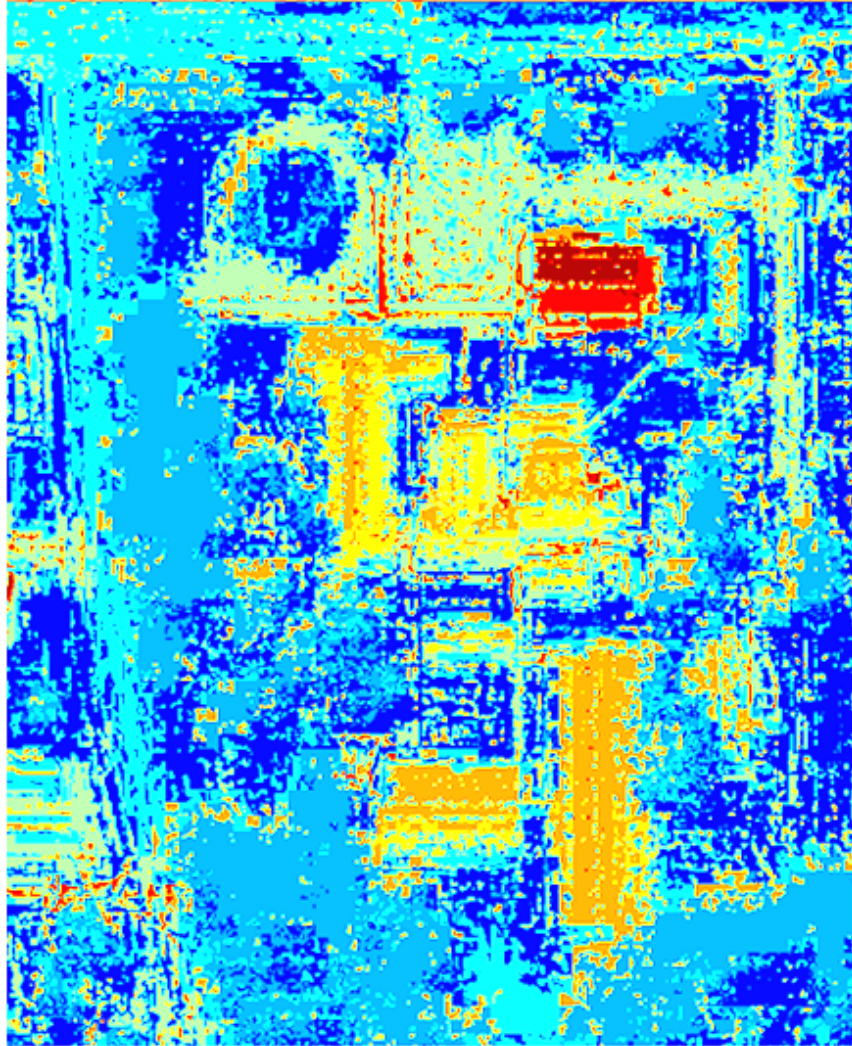
Figure 31. Material map showing SAM results constrained to HSI linear least squares unmixing abundances.

In cases where a relatively small number of endmembers are used to describe a

scene, materials are likely to be continuous over small spatial areas. Using this

assumption, morphology techniques are applied to smooth the classification image. An

opening followed by a closing using a small structure element disk of radius three (which

is actually a three by three square due to the limited number of included pixels) reduces

some of the noise caused by probable incorrect assignments in an otherwise continuous region. The result after smoothing is shown in Figure 32 below along with the HR image for comparison and a magnified material map resulting from performing SAM using the HSI spectra of the same endmembers and only the original HSI cube. Note that by using only the HSI information, SAM has trouble distinguishing road from the asphalt shingles of the roofs and parking lot from asphalt.
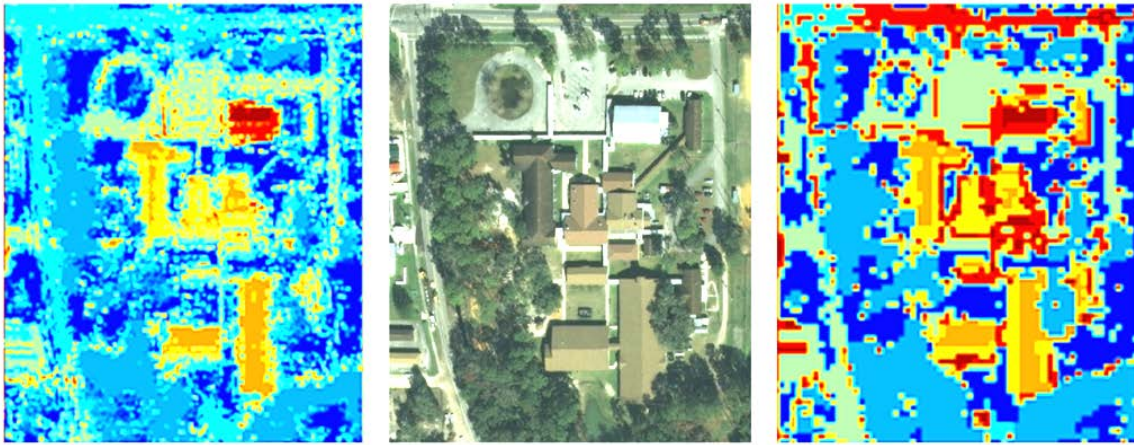
Figure 32. Left to right: The resulting material map after performance of morphological opening followed by closing, HR image shown for comparison, SAM classification performed on the original HSI for comparison.

While this linear unmixing fused method of Approach 2 does provide an improvement in edge information over the material map resulting from the application of SAM on the HSI alone, the HR material map still appears to include an unsatisfactorily high number of incorrectly identified pixels. This is likely due to the imperfect endmember spectra, even though they are selected by hand, and the strict enforcement of the determined abundances using those endmembers when performing pixel assignment at the higher resolution. While further efforts could be employed to attempt to improve this method, the next fusion approach provides better results with less effort.

4.8   Fused ISODATA Material Map Extraction: Approach 3

The final and most successful approach for automating the generation of a material map using all of the information available begins with the HR and rasterized lidar elevation and return strength to perform image segmentation before incorporating the spectral information of the HSI (Givens, et al., 2013). Generally, the RGB imagery has the highest resolution of the three modalities. However, in many cases the RGB imagery has shadowing and thus classification of the RGB imagery alone can result in a shadow class which is not a proper scene material. With the desired output being a material map that can be used to generate a three-dimensional model of the scene, the presence of a shadow class creates areas of shade in the scene regardless of how the scene is illuminated in the model, and thus reduces its utility for modeling illumination angles other than the one under which the original RGB is captured. To avoid a shadow class in the material map, Approach 3 fuses the RGB imagery with the rasterized lidar return strength and elevation which lacks shadows due to its active illumination collection method.

As in Approach 2, the result is a five-band cube with the first three bands being the red, green, and blue from the RGB image, the fourth being the rasterized lidar return strength, and the fifth being the rasterized lidar elevation as measured by the first return. To create this fused cube, the rasterized lidar bands are transformed and resampled to the resolution of the RGB image using the transformation matrix obtained from the registration method. The shadowing and illumination angle effects in the RGB bands are further reduced by normalizing out the intensity at each pixel using Equation (4.1):

$$R_{norm} = \frac{R}{\sqrt{R^2 + G^2 + B^2}},$$

$$G_{norm} = \frac{G}{\sqrt{R^2 + G^2 + B^2}}, \qquad (4.1)$$

$$B_{norm} = \frac{B}{\sqrt{R^2 + G^2 + B^2}},$$

where R, G, and B are the original red, green, and blue values at each pixel and R$_{norm}$,

G$_{norm}$, and B$_{norm}$ are the new normalized values at each pixel. The final consideration

before performing image classification on the fused RGB and lidar cube is weighting of

the respective bands. Because a shadow class is not desired and shadowing is still

apparent, though reduced, in the color-normalized RGB image, the rasterized lidar return

strength band is weighted higher than the color-normalized RGB bands. The lidar

elevation is weighted lower than the lidar return strength because, although elevation

does help to discriminate between ground and not ground, materials that are not ground

which have varying elevation (like angled roofs and tree canopies) are broken into

different classes if the elevation weighting is too high.

Preliminary image classification is then performed on the five-band cube using

the common linear classifier, Iterative Self-Organizing Data analysis technique

(ISODATA) (Eismann, 2012). ISODATA performs image classification by assigning all

pixels in the cube to a set of randomly selected initial class vectors using the nearest

mean classification rule (minimum Euclidean distance) shown in Equation (4.2):

$$d_q(x) = (x - \mu_q)^T (x - \mu_q) = \|x - \mu_q\|^2, \qquad (4.2)$$

where d$_q$(**x**) is the calculated distance of spectrum **x** from the mean spectrum of class q,

μ_q. Class means are then updated using the new class populations and the algorithm

iterates from there (Eismann, 2012). A major advantage of ISODATA is that the initial

number of classes specified by the user can be reduced by the algorithm if a class

population becomes too small or increased if the standard deviation of a class becomes

too large. Thus, only an approximate guess of the number of classes in a scene is

necessary. As in Approach 1 and Approach 2, the initial guess for number of classes is

obtained by performing a principal components analysis on the HSI and using the linear

mixing model rule of thumb in which one assumes the number of classes in the scene is

one greater than the number of significant PCA bands where the number of significant

PCA bands is determined by setting a threshold on total variance captured (Eismann,

2012). ISODATA has a potential drawback in that it requires six user-defined parameters

and thresholds besides the initial number of desired classes, which could alter the

algorithms effectiveness if these values are not adjusted for different scenes. While other

unsupervised clustering algorithms with fewer user inputs could be implemented (the

Improved Split-and-Merge Clustering algorithm for example), ISODATA provides good

results on the datasets used in this research with minimal deviation from the six default

parameters. The end result of running the ISODATA algorithm on the fused RGB and

lidar cube is a material map at approximately the resolution of the HR imagery where

pixels of similar color-normalized red, green, and blue intensity; lidar return strength

intensity; and lidar elevation are classified as the same materials.

The ISODATA material map is smoothed using an adaptive median filter to

eliminate apparent noisy assignments and further smoothed by requiring a minimum

number of pixels in assigned regions (Givens, et al., 2012). In addition, the ISODATA

results tend to have thin borders of likely misclassified pixels around large uniform

regions of a single class. The thin borders are reduced by performing a morphological opening of each of the classes, then iteratively dilating each of the classes to fill the voids left by the openings. While each of these smoothing steps might eliminate small regions of correctly assigned pixels, the low-resolution HSI cannot provide pure pixels to further investigate them anyway. These smoothing steps introduce additional user inputs which may need to be adjusted depending on the scene under investigation.

The next step in this classification process is to check whether pixel regions assigned to a class by the ISODATA algorithm have similar spectral content in the HSI. However, to be able to compare spectra of different regions, the regions must be of large enough area to ensure they cover a full HSI pixel. To ensure the regions are large enough, the smoothed ISODATA result is transformed to the HSI resolution and then borders of continuous regions are eroded by two pixels. Using connected components, the spectral average of the remaining pixels of each region are then compared to other regions of the same class label by determining the spectral angle between the two. If the spectral angle falls above a threshold, the region is labeled as a new class. If not, the class average spectrum is updated to include both regions. Unfortunately, in its current state this step can only be applied to pure pixels, so regions of the ISODATA material map that are too small to ensure a pure HSI pixel are not included in this step. The results of the HSI spectral comparison are then transformed back to the HR space of the five-band cube and logic steps are applied to separate classes of different materials and potentially merge classes that are spectrally similar. The merging of similar classes is not especially important for synthetic scene generation since there is no limit on the number of material classes used in DIRSIG. Thus, it can be bypassed to eliminate the chance for

it to introduce errors. The flowchart in Figure 33 shows the main steps involved in

Approach 3 along with the additional bypass arrow for the merging logic step.
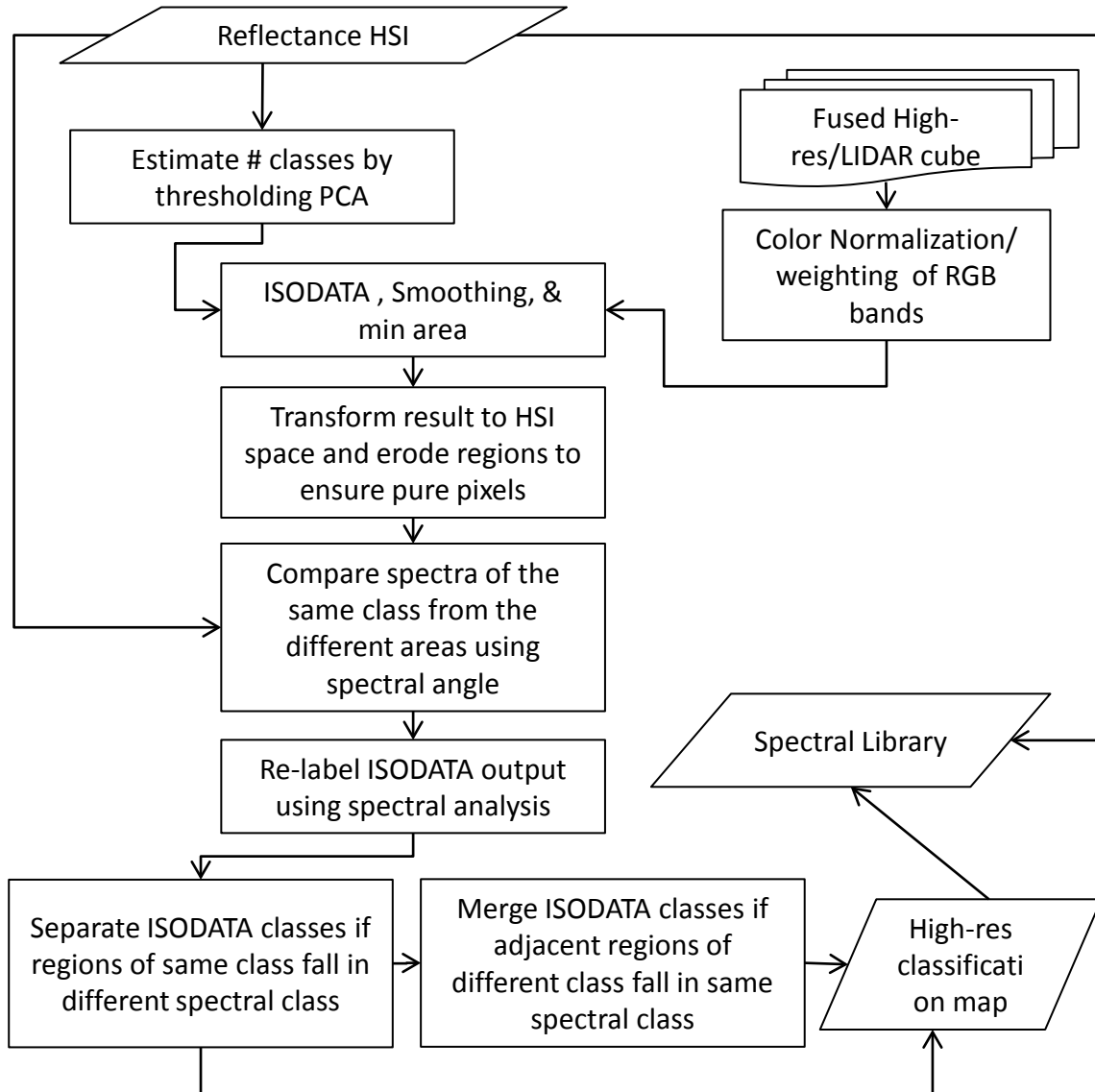
Figure 33. Flowchart of the main steps of Approach 3.

4.8.1   Approach 3 Walkthrough on the Reduced ROI Melrose Scene

For easy comparison to Approach 2, the same cropped Melrose ROI is presented

to show the results of each step in Approach 3. As stated in the Approach 2 section, the

region includes a mix of vegetation and manmade structures as well as elevation variation

and is therefore a good challenge for the classification routine. Once cut down to approximately the same ROI, the datasets are registered using the Walli registration method. The resulting RGB is then color normalized using Equation (4.1) to minimize the shading in the image. Both images are shown below in Figure 34 and the rasterized lidar return strength and elevation are shown in Figure 35.
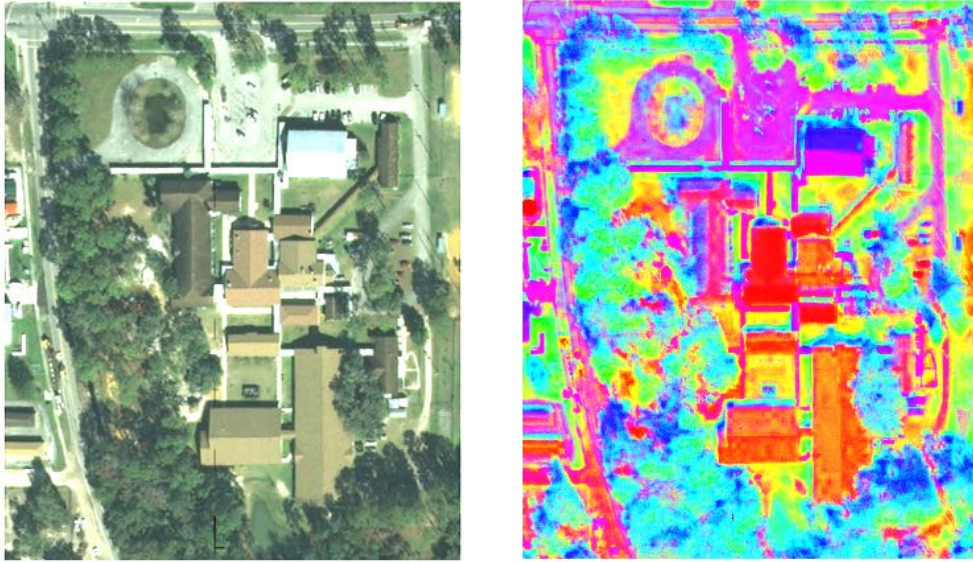
Figure 34. Left: The RGB of region of interest, Right: The color-normalized RGB used as the first three bands of the fused data cube for ISODATA (histogram equalized here for viewing).

The concatenation of the weighted, color-normalized RGB and the two rasterized lidar bands forms the cube that is input into the ISODATA algorithm. The ISODATA result with a user input of seven desired classes is shown in Figure 36 below. This result also includes an application of adaptive median filtering to reduce salt and pepper type incorrect assignments along with morphological steps to enforce minimum region size and eliminate thin border regions.
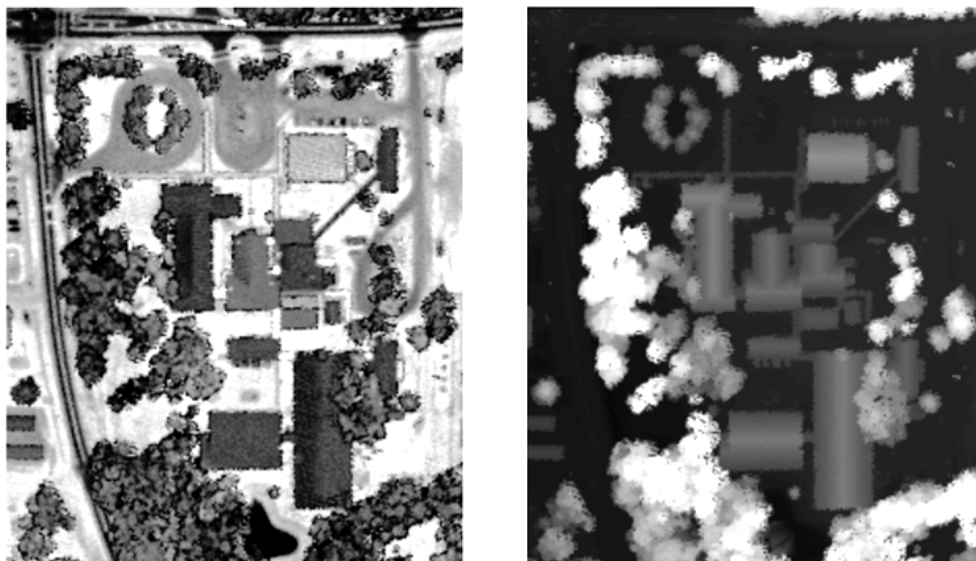
Figure 35. Left: The rasterized lidar return strength of the region of interest used as band 4 in the data cube input to ISODATA, Right: The rasterized lidar elevation of the region of interest used as band 5.
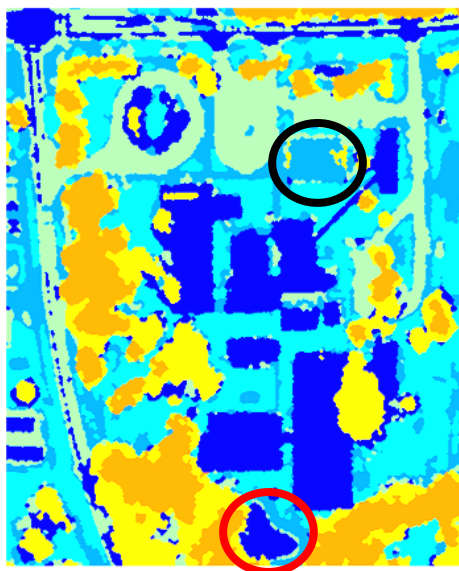
Figure 36. The ISODATA output of the fused, five band cube after smoothing and minimum region size. Initial number of bands specified is seven, but ISODATA finished with six.

As these results show, ISODATA does an impressive job of segmenting the cube, but

some class confusion exists. For example, note that the non-rectangular dark blue region

at the bottom of the image (circled in red) falls in the same class as the asphalt shingle

78

roofing of the buildings and asphalt road, but is actually a small pond. Also, the white, metal roofed building class (light blue, circled in black) falls in one of the grass classes. Finally, assuming they are spectrally similar, the orange and yellow classes represent classes that should probably be merged as a single tree class but were separated by ISODATA due to the varying height of the tree canopy.

The HSI data is incorporated into this process by applying logical rule-based classification. First, the smoothed ISODATA material map is transformed to HSI space and connected components are eroded by two full HSI pixels using MATLAB's imerode function with a two pixel radius disk. The ISODATA result at HSI resolution and resulting map after erosion to pure pixels are shown in Figure 37 below.
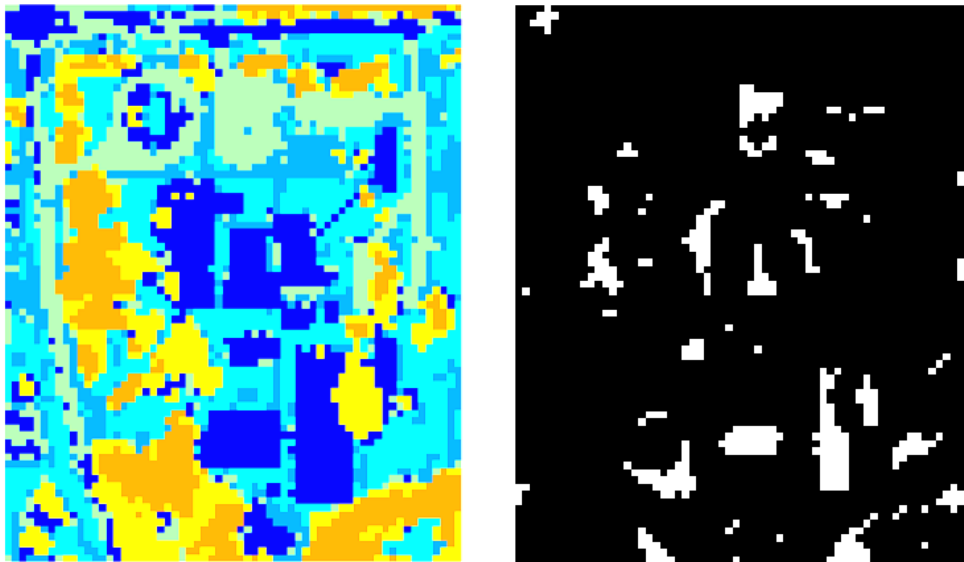
Figure 37. Left: The ISODATA result transformed to HSI space, Right: The remaining connected components after erosion of outer two pixels.

Next, spectral angles between mean spectra for each of the connected components are computed. If the spectral angle between two connected components falls over a threshold (0.15 radians), a new class is created. Otherwise, the two regions are averaged

together to create a combined class average spectrum. The next connected component is then compared to the first two and so on. Figure 38 shows the resulting classification of the pure pixel connected components in which five distinct materials are determined to be present and where each color refers to a separate class.
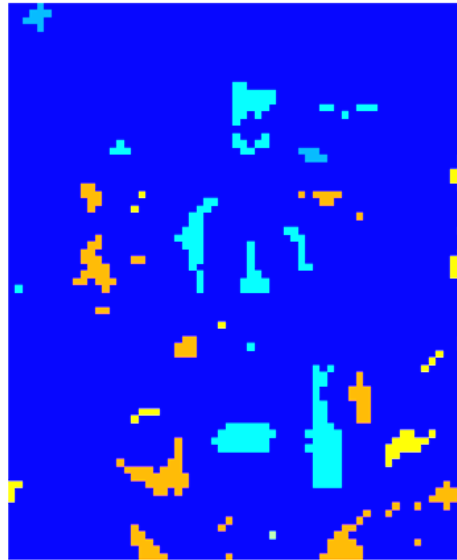
Figure 38. The resulting material map using spectral angle measurement between connected components.

This eroded class map from the HSI is then transformed back to the HR space for the application of logic rules. The first rule is that connected components which share the same ISODATA result, but receive different HSI classes are labeled as different classes. In the event that two different HSI results occurred in the same ISODATA connected component (i.e. a connected component in the HR space that, when transformed to HSI space and eroded, become two or more connected components which then received separate spectral angle classifications) the spectral angle result assigned to the majority of the pixels in the connected component region is used for the entire connected component. The resulting material map after applying this rule is shown below in Figure 39. Notice

that the pond is separated from the rooftops and road asphalt classes by this step and the

white roofed building and covered walkway is separated from the grass class. However,

the trees mislabeled as asphalt roof top in the center of the circular drive are still

incorrectly labeled because these regions are eroded away when transformed to HSI

space and thus deemed too small to include pure pixels. A different type of rule is

needed for small regions and will hopefully become available through future research.

Additionally, the application of this rule using spectral angle separates the grass class into

three separate classes and the asphalt shingle from the asphalt road class. Changing the

minimum spectral angle for class separation would alter this result if less or more classes

are desired.
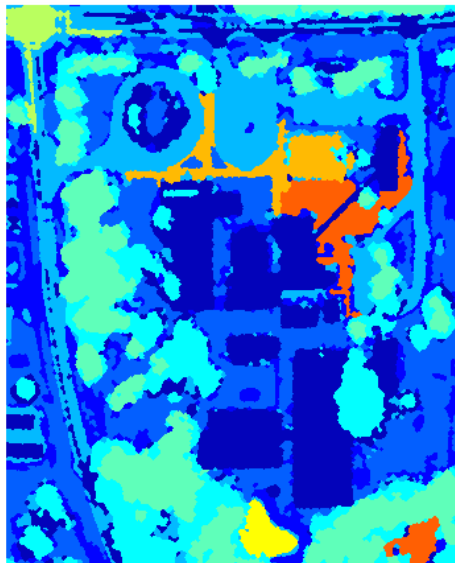
Figure 39. The resulting material map after applying first logic rule using the HSI spectral angle results.

Next, connected components which are labeled as different classes by ISODATA,

but are identified as the same class by the spectral angle comparison are labeled the same,

but only if they are adjacent to each other. This step could be changed to label all

connected components with the same spectral angle as the same regardless of their spatial

relationship to each other. For this scene eliminating the adjacency requirement results in

a combined class for some of the roof tops and road asphalt. The user's end goals for the

classification dictates whether this result is preferable. Enforcing the adjacency

requirement, this step merged some of the adjacent tree and grass regions into single,

respective tree and grass classes. Below in Figure 40 is the final result compared to the

initial ISODATA result from Figure 36 and a material map produced by using the

unsupervised Stochastic Expectation Maximization algorithm on the HSI alone for

comparison.
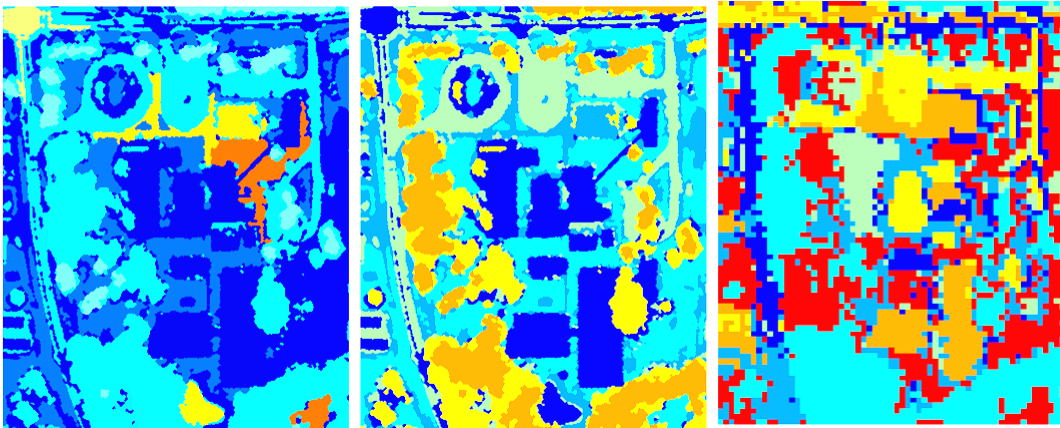
Figure 40. Left to right: The final resulting material map from Figure 39, the initial ISODATA segmentation of the fused RGB/lidar cube, and an example SEM classification using only the HSI for comparison.

As the results in Figure 40 show, using the ISODATA algorithm on a fused, HR cube

with proper weighting of the bands followed by logic based rules to incorporate spectral

information from the HSI results in a material map with approximately ten-fold increase

in resolution compared to using the HSI alone.

The following example shows another case in which the erosion step, which

ensures the comparison of only pure pixels at the HSI resolution, can produce incorrect

results. In this example, a slightly larger region of interest from the Melrose scene is used. Figure 41 shows the ISODATA results on the left with a green circle identifying where the error occurs. This is the same bright white rooftop which is identified in the previous example and successfully separated into its own class. In this case, however, the initial segmentation by the ISODATA results in a the rooftop being classified into two separate classes likely due to the difference in illumination angle and therefore brightness of the two halves of the roof. Since the two classes are fairly narrow in the vertical direction, when the ISODATA result is transformed to the HSI space (the middle image in Figure 41) and then eroded to ensure pure pixels (the right image in Figure 41), the two regions are eroded away and no spectral comparison is performed. Thus the two roof halves retain their initial ISODATA assignment which has the rooftop split between a tree and asphalt class.
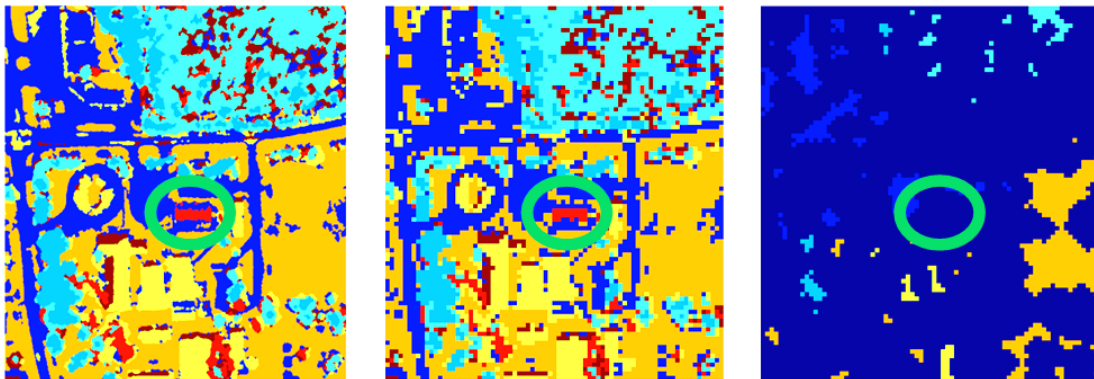
Figure 41. Additional example of Approach 3 showing how small regions at the high resolution can be eroded away and not included in the spectral comparison potentially leading to incorrect results.

Even with the possibility of incorrect classification of small pixel regions, the test results from Approach 3 show a large improvement in resolution of the material maps. Thus, the approach is included as another fully automated process for producing a

material map in the LHD method and presented here on the full VanLare and Melrose datasets. Again, the VanLare dataset is a synthetic dataset created using DIRSIG with a high-fidelity scene. The Melrose dataset is a real dataset collected over a site near Melrose, Florida for the NEON project. This section shows the steps and results of the LHD method using the ISODATA fusion method of Approach 3 to generate a HR material map.

4.8.2 Approach 3 Applied to the VanLare and Full Melrose Datasets

The first step of the Approach 3 HR classification process is to run the ISODATA classification algorithm on a five-band, fused cube consisting of the red, green, and blue bands of the HR imagery and the rasterized return strength and elevation from the lidar data. It should be noted that, for this case in which the inputs for the method are DIRSIG-generated, the rasterization of the lidar return strength produced by DIRSIG is good enough to perform the registration, but too noisy to produce consistent classification results. DIRSIG uses a photon mapping technique to produce the lidar returns so noise in the lidar returns could be reduced by including more photons in the simulation, but this caused run times to become prohibitively long. Instead, to generate a return strength image consistent with the images obtained using real lidar instruments, a passive imager centered at 1064 nm was placed over the scene in DIRSIG. Additionally, instead of using the sun as the illumination source, which would have produced shading in the scene and would therefore not have been representative of the actively illuminated return strength obtained using a lidar instrument, illumination was provided by modeling the sky as a 5000 K blackbody instead of its more common 260 K temperature. This created the return strength image seen as band 4 in Figure 42.

84
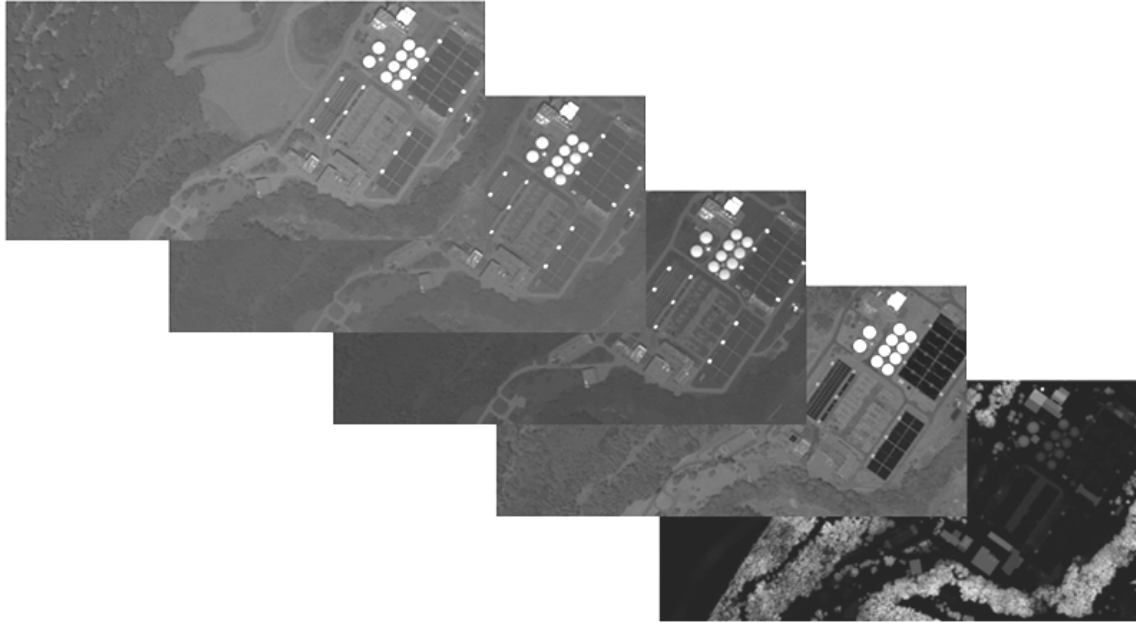
Figure 42. The five band cube used to perform image segmentation with ISODATA. Bands one, two, and three are the color-normalized red, green, and blue bands from the HR imagery, band four is the modeled IR return strength, and band five is the rasterized lidar elevation.

ISODATA is then run on the five band cube with the initial condition to classify the pixels into seven classes. When applied to the VanLare scene, the principal component dimensionality reduction results in six significant PCA bands and therefore seven classes are used to initiate the ISODATA algorithm. The number of significant principal component bands determined in the Melrose data case is 15 and is too computationally demanding for the home desktop computer used for the processing. Since the number of significant principal component bands can be adjusted simply by adjusting the threshold amount of variance to capture, the number of significant bands is somewhat arbitrary and seven classes seemed to produce good results using the ISODATA algorithm and this five-band cube. The ISODATA results are shown in Figure 43. Some slight smoothing and morphology are employed to eliminate very small

connected components and some thin line artifacts that tend to be produced around building edges.
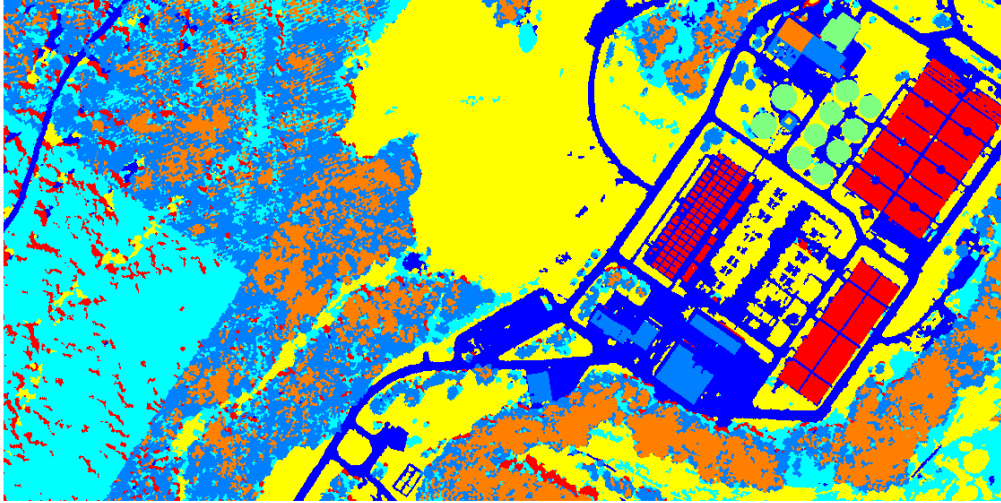
Figure 43. The initial ISODATA classification results using the fused five-band cube consisting of the three RGB bands from the HR imagery, a modeled IR return strength band, and the rasterized lidar elevation band.

After the initial image segmentation by the ISODATA algorithm, the next step is to further separate the classes by relying on the spectral component given in the HSI. However, since the goal is to improve the classification resolution, the connected component edge information gained from the fused HR and lidar image is valued higher than the hyperspectral component. In other words, the segmentation resulting from the HR ISODATA step is not changed by the information gained from the spectral HSI component. Only the material class assignments of entire connected components are adjusted. Thus, when the segmented image is transformed to hyperspectral space, only the pure pixels are retained. The pure pixel map is generated, by eroding the two outermost layers of pixels of each of the connected components in the HSI-space segmented image. The simple spectral angle comparison algorithm is then run to

86

differentiate the classes. It starts with the upper-left-most connected component in the scene and sequentially steps through all of the other connected components. If the next connected component differed spectrally from the first by more than 0.15 radians, it is considered a new class. If not, the spectra are averaged and the next connected component is evaluated. The result of that process is shown in Figure 44.
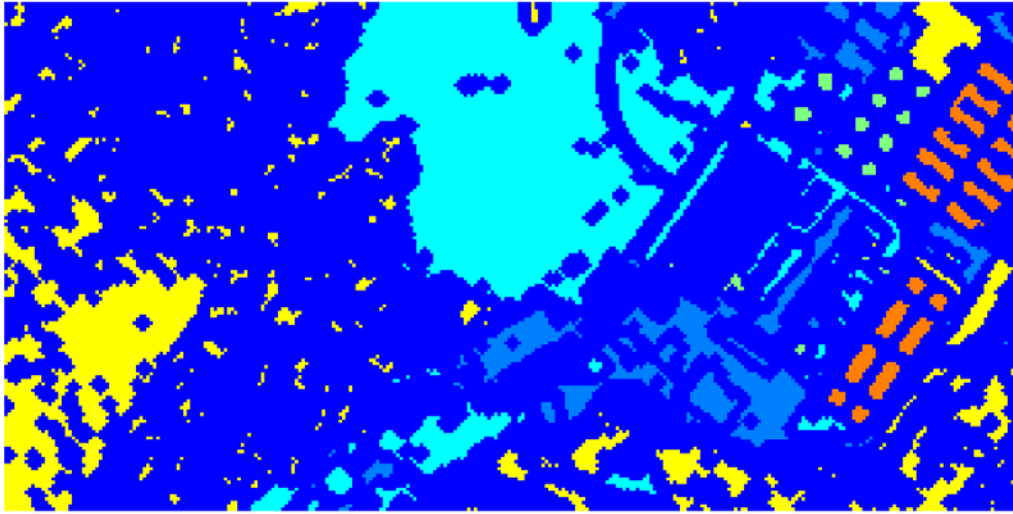
Figure 44. The resulting material map after a thresholded spectral angle is computed for the eroded connected components.

Though somewhat difficult to distinguish in the Figure 43 image, the ISODATA step classifies roof tops and a large portion of the trees as the same material. In the SAM image in Figure 44, these two classes are correctly identified as different materials. By taking this new spectral segmentation back to the HR space and applying the logic step that connected components identified as spectrally different must be labeled as separate classes, the resulting final HR material map shown in Figure 45 is obtained. This material map is then transformed to the lidar coordinates and draped over the lidar using the same technique as in the texture maps from the previous section.
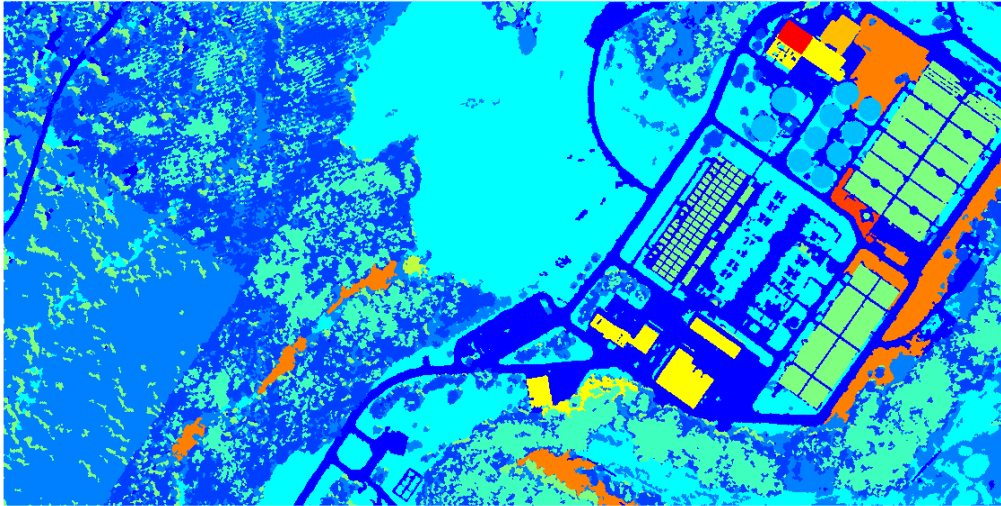
Figure 45. The final HR material map after the HSI logic step is applied.

Stepping through the same process on the Melrose dataset produces the material

map shown in Figure 46 below.
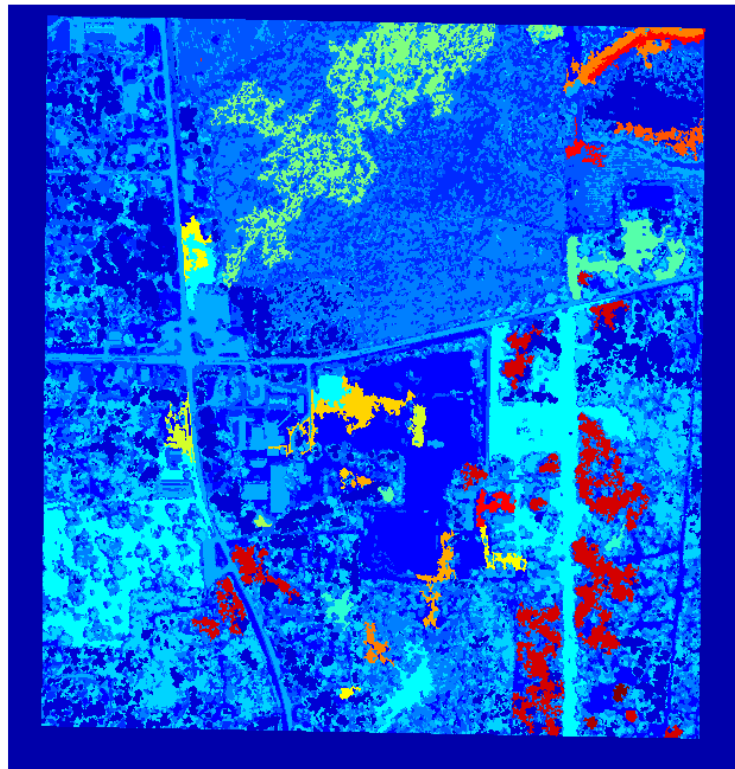
Figure 46. Top: The resulting HR material map produced using the ISODATA fusion method on the Melrose dataset.

4.9   Extraction of the Material Library

For the final step of the LHD method, a reflectance file for each material class is generated using the material map (obtained through either Approach 1 or 3) to select up to 300 reflectance curves (pixels from the HSI) from each of the full classes. It is important to note that for both the VanLare and Melrose datasets, the HSI is in reflectance units. For the VanLare case, reflectance is achieved simply by dividing by capturing the radiance of a perfect spectralon tile placed in the synthetic scene. For the Melrose dataset, the HSI is provided in reflectance units with atmospheric compensation accomplished using the FLAASH algorithm (Krause and Kuester, 2011). If the Approach 3 material map is used to generate the spectral library, it must first be transformed to HSI space and eroded to ensure selection of pure pixels.

The reflectance curves are represented as an emissivity in the DIRSIG spectral library which is equal to one minus reflectance for opaque objects under the assumptions of Kirchhoff's Law (using Equation (3.9)). The selection is performed using a uniform random selection with respect to pixel indices ensuring the class statistics of each selected subset reflect the class statistics of their respective full classes. Lambertian materials are assumed for simplicity. The scene specific emissivities for each material are then written to a text file consisting of pairs of column vectors for wavelength and emissivity for each spectrum in the class as DIRSIG requires. The final HR material map is also used to locate each class in the red, green, and blue texture maps and generate the material means and standard deviations for each map. Figure 47 shows the full Mahalanobis distance statistics for a vegetation class identified using the SEM method of Approach 1 and the corresponding statistics of the 300 randomly selected reflectances

pulled from that class to be written to the emissivity file. The plots show that the shape

of the Mahalanobis distance distribution of the subset is similar to that of the full set and

thus the uniform random selection is successful in capturing a representative subset of

spectra for the class.
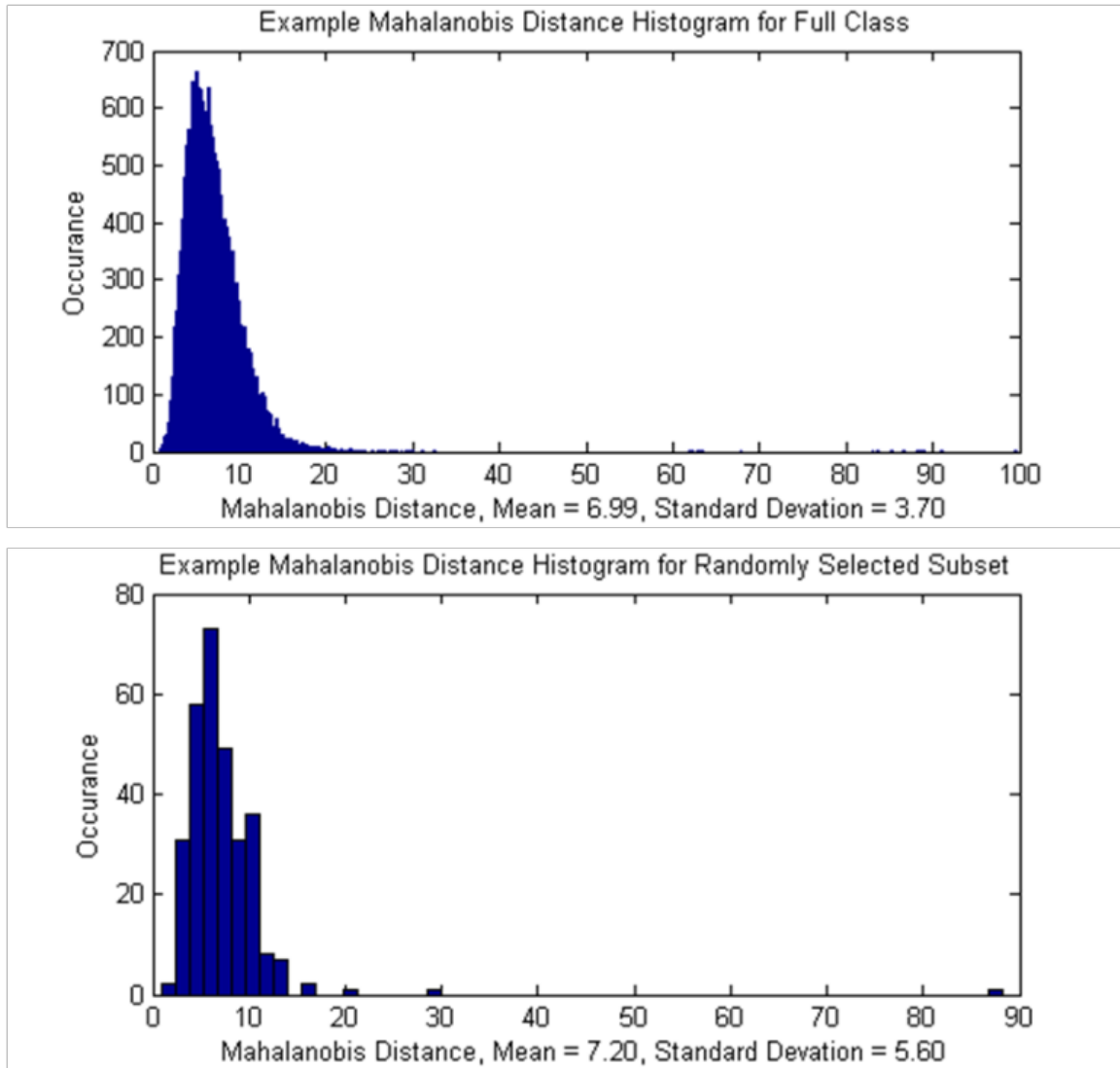
Figure 47. Top: Example class statistics for one of the SEM determined endmember classes. Bottom: Statistics for uniform random selection of 300 spectra ensuring the emissivity file contains a range of spectra representative of the entire class.

4.10   Chapter Summary

This chapter presents the end-to-end process for automating the extraction of a synthetic scene from HR imagery, lidar, and HSI using the LHD method. It begins with an introduction to the VanLare and Melrose datasets, then includes explanations and example results for each step in the registration process, geometry extraction, and texture map extraction. Next, the three new approaches to automated material map extraction are presented. Approach 1 is an automated, non-fused approach, Approach 2 is an attempted fused approach which is not successfully fully automated and produced less than satisfactory results, and Approach 3 is a fused, fully-automated approach. The three approaches are presented in detail along with flowcharts showing the main steps in each approach and example outputs for each step. The final section of the chapter discusses how material libraries are autonomously extracted from the HSI using the material maps generated from Approaches 1 and 3. The next chapter presents DIRSIG-generated synthetic imagery using the autonomously generated scenes of this chapter and compares the quality of the results from Approaches 1 and 3 to the original scene inputs.

V. Comparison of Scene Construction Methods

5.1   Chapter Overview

This chapter evaluates the quality of DIRSIG-generated synthetic imagery of the VanLare and Melrose scenes built using Approaches 1 and 3 in the LHD method. Both RGB imagery and HSI are presented for a qualitative evaluation of the Approach 1 and 3 results when compared to the original input RGB imagery and HSI. Additionally, automated registration between the synthetic imagery from the LHD scenes to their respective scene inputs is shown to provide confidence that the recreated scenes are accurate enough to perform multimodal registration. While the qualitative evaluation and automated registration test provide good confidence in the accuracy of the synthetic scenes from both the non-fused SEM Approach 1 and the fused ISODATA Approach 3, evaluation of actual scene accuracy is difficult without a quantitative comparison to the original input imagery. Thus, this chapter also presents a quantitative comparison using an average band difference metric and spectral angle metric. These comparison metrics are first applied to the VanLare dataset followed by the Melrose dataset. Finally, while the average band difference metric shows that both Approach 1 and Approach 3 produce accurate synthetic scenes, it fails to quantitatively capture the sharpening improvements made by Approach 3. Thus an additional metric is included to help capture this result and present ideas for future research in the evaluation of synthetic scene accuracy.

5.2   VanLare Scene Comparisons

By generating a synthetic scene from the synthetically generated VanLare dataset, the DIRSIG outputs of the recreated scene are easily compared to the original

DIRSIG generated inputs because the camera positions and designs, atmospheric conditions, and illumination conditions are all duplicated exactly. Thus no registration or other matching is required between the input and output images to perform qualitative and quantitative comparisons. Before performing quantitative comparisons, the first and easiest comparison is a qualitative examination of the RGB imagery. Here, the exact same sensor and placement is used over both the original high-fidelity scene and the new recreated scenes. The high-fidelity input RGB is shown in Figure 48 above the synthetic RGB imagery created using the recreated scenes of Approach 1 and Approach 3 in DIRSIG. While the three images in Figure 48 look very similar, under close inspection the middle image, generated using the lower resolution material map of Approach 1, shows pixelation around many of the thin border regions especially in the water holding ponds on the right side of the image. Otherwise, the qualitative comparisons of Approach 1 and Approach 3 show little difference.

To show that the recreated model is representative of the original scene, synthetic imagery generated using the recreated model can be registered to the input 'original' synthetic imagery. The successful quarter-pixel registration of the input RGB and recreated RGB from Approach 1 shown in Figure 49 provides an additional check that the recreated scene is a good representative of the original input scene. The resulting transformation matrix and match information are provided in Table 6.

Figure 48. Top: The DIRSIG generated RGB image over the "original" high-fidelity scene at 0.3 meter GSD, Middle: The DIRSIG generated RGB image using the non-fused SEM Approach 1 material map, Bottom: The DIRSIG generated RGB image using the ISODATA fused Approach 3 material map.

Figure 49. Registration of DIRSIG RGB image and real RGB image of the VanLare scene. The green lines indicate the 692 'good' SIFT matches which survived culling by both RANSAC and RMSDE minimization.

Table 6. The registration results of the original input RGB image to the RGB image produced using the recreated model.

High res RGB Original to DIRSIG Registration		
1.0016	0.0004	0
0.0002	1.0020	0
-1.8290	-1.1477	1
RMSDE Average: 0.2495 pixels		
Matches Remaining: 692		

For a more quantitative comparison, an average intensity difference per band is calculated to compare the RGB images from the recreated scenes to the original input RGB images. The calculation is performed at each pixel using the following equation

$$D_{ij} = \frac{\sum_n \left| x_{n_{i,j}} - u_{n_{i,j}} \right|}{n},$$
(5.1)

where D_{ij} is the average difference per band and i and j specify the pixel location, n is the number of bands in the image and **x** and **u** are the spectra (or RGB values in the case of RGB images) of the original and recreated scenes. Figure 50 below shows the resulting images using this method of comparison. While both scenes produce relatively low average error of 4.5 digital counts per band (on a scale of 0 to 255) the pixelation error around the water holding pools is readily apparent in the top image comparing the scene using the low-resolution material map of Approach 1 to the original RGB. However, the

95

bottom image shows larger error along the tree line in the bottom right portion of the image where the ISODATA fusion classification of Approach 3 appears to have misclassified some grass areas as tree. As detailed in the previous chapter, this thin region is another instance in which the connected components are too small to include pure pixels for spectral comparison when the HR ISODATA result is converted to HSI space and eroded. For both cases, the largest variations are seen around the edges of the buildings where there could be slight registration error in addition to variations in the geometry due to the coarser sampling of the lidar.
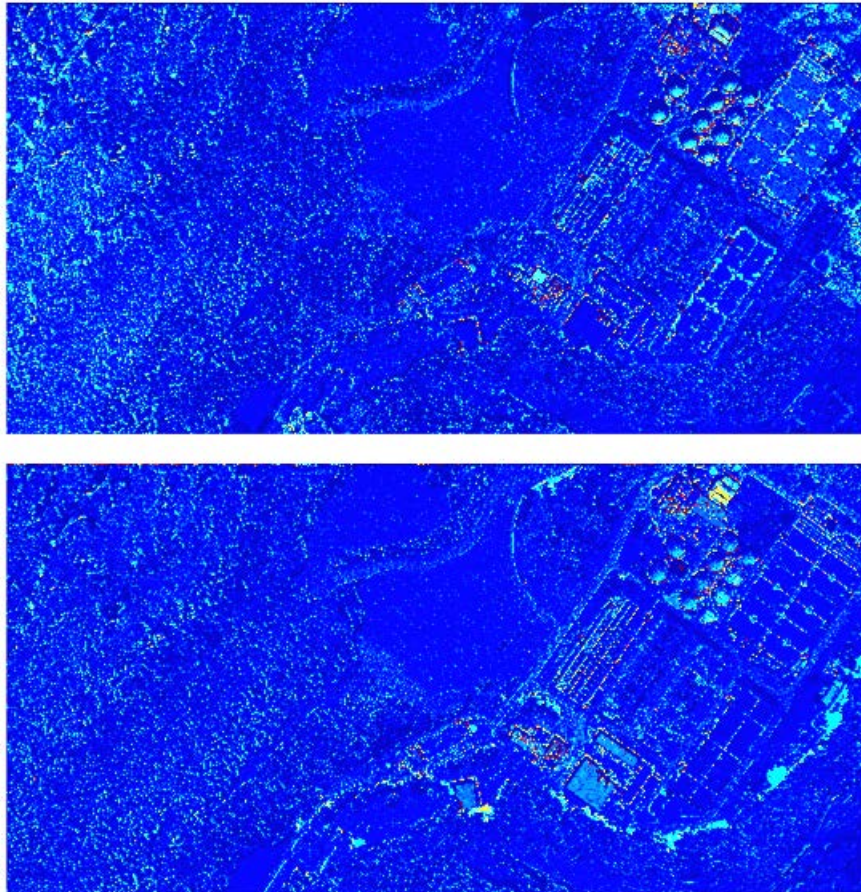
Figure 50. Top: The average intensity difference per band image comparing the low-resolution VanLare scene to the original, Bottom: The average intensity difference per band comparing the HR VanLare scene to the original.

Similarly, the HSI can be compared both qualitatively and quantitatively. Figure 51 visually compares the HSI of the original high-fidelity scene to the low-resolution Approach 1 and HR Approach 3 recreated scenes. These RGB images are created using the 650 nm, 550 nm, and 450 nm bands of the respective HSI outputs.

The full content of the HSI can be compared using the same average band difference calculation from Equation (5.1). Here the actual value has a more physical meaning since the compared cubes are in units of percent reflectance. Therefore, each pixel in the difference images are the average reflectance difference per band at that location. The image-wide average band difference for the low-resolution Approach 1 scene to the original scene comparison is 2.0% average reflectance difference per band and the average band difference for the HR Approach 3 scene is 2.2% average reflectance difference per band. As in the RGB comparison, the largest differences are in the tree regions and around building edges as can be seen in Figure 52.

While the difference images give an indication of intensity difference between the two HSI outputs, a better spectral comparison can be accomplished by computing the spectral angle between the high-fidelity and recreated HSI cubes at each pixel. These images are shown in Figure 53. Here the angles in radians have less physical value since they are an angle in the n-dimensional space of the hyperspectral cubes (210 bands in this case), but the image mean is 0.05 radians for both cases. It can be noted, however, that many of the larger angles tend to occur along shaded regions for both cases, but overall the recreated scenes are in better agreement spectrally when the intensity component is not included.
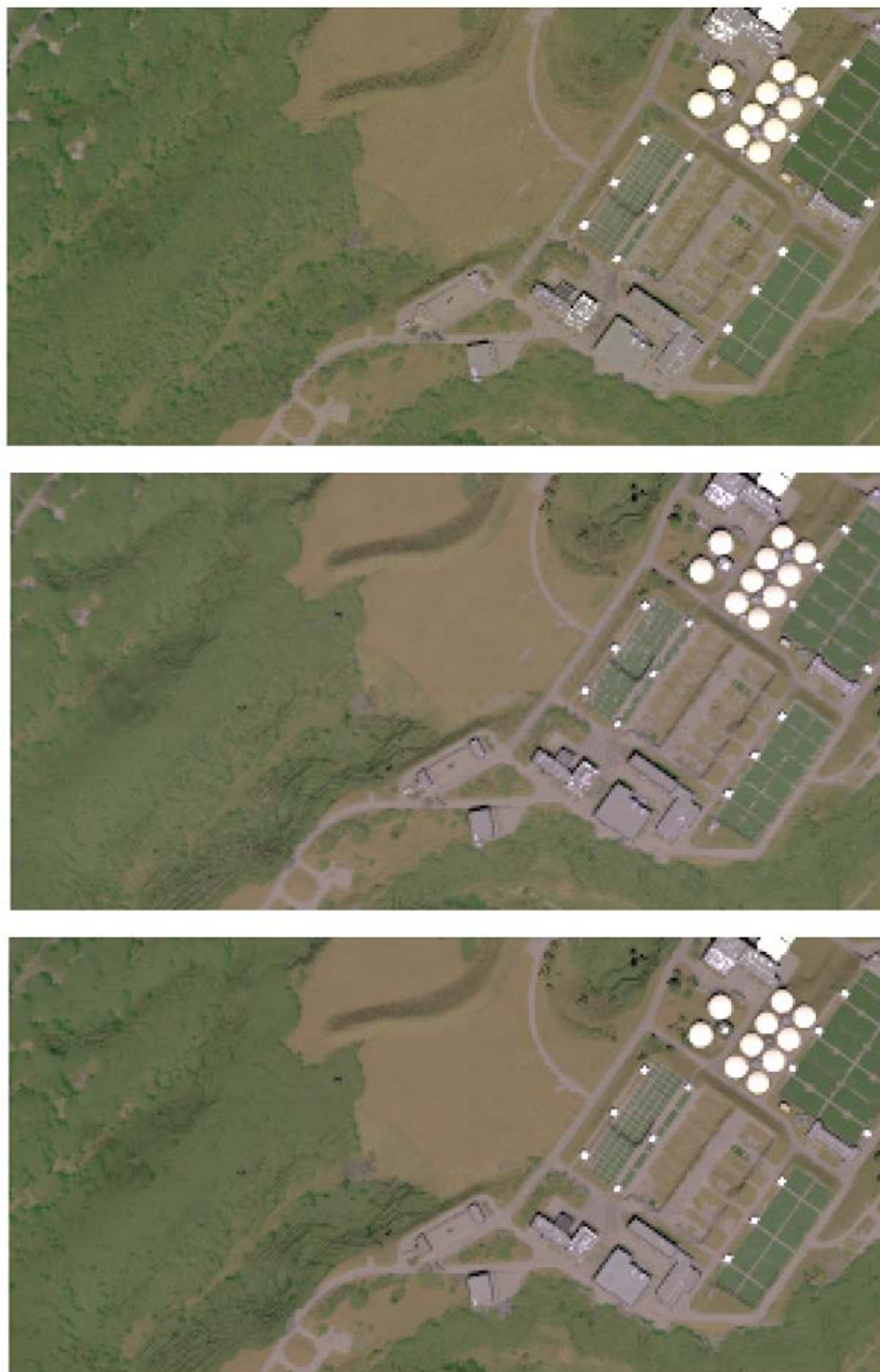
Figure 51. Top: Color composite of the original HSI, Middle: Color composite of the low-resolution Approach 1 recreated HSI, Bottom: Color composite of the HR Approach 3 recreated scene.
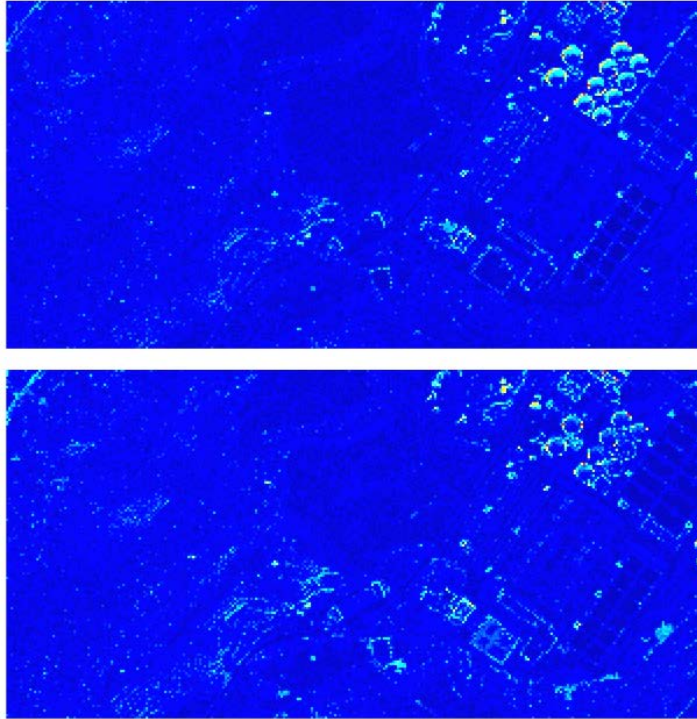
Figure 52. The average reflectance difference per band for the HSI of the low-resolution scene (top) and HR scene (bottom).
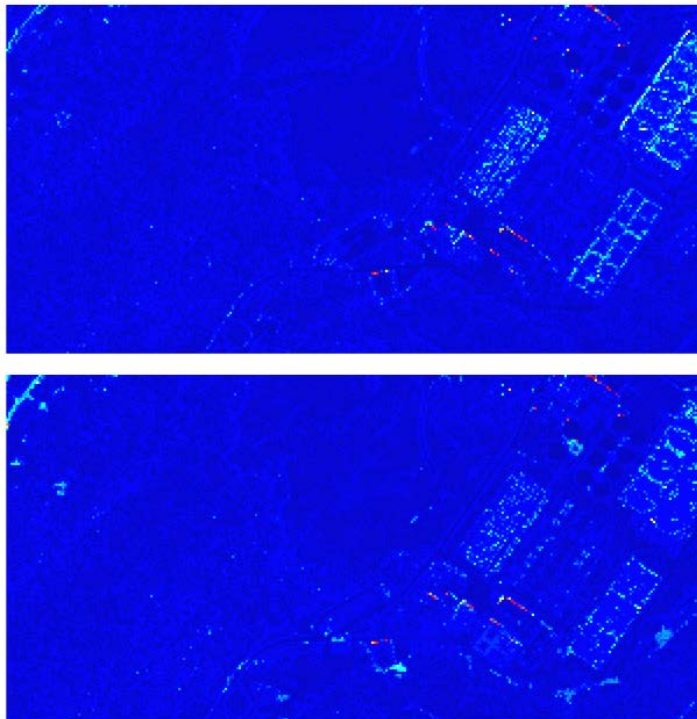
Figure 53. The spectral angle image comparing the HSI of the original high-fidelity scene to the HSI of the recreated scene.

5.3   Melrose Scene Comparisons

In this section, the same comparisons are performed on the real Melrose dataset as in the VanLare scene case. However, since exact camera parameters, date and time of collect, and atmospheric conditions of the input imagery are not known in this case, these parameters are estimated when using DIRSIG to produce the synthetic RGB and hyperspectral imagery for comparison. Additionally, since camera locations are not known, a registration of the synthetic imagery to the input imagery is necessary before quantitative comparisons can be made. This extra step introduces additional error in contrast to the fully modeled VanLare case in which no registration is required since camera positions and parameters are reproduced exactly between the original and recreated images.

For the qualitative comparison, though some pixelation is again evident in the low-resolution Approach 1 recreated scene, the RGB imagery from both the low-resolution and HR Approach 3 Melrose scene compare well to the input FLDOT RGB imagery as seen in Figure 54 below.

The registration of the Approach 1 RGB to the input RGB is shown below. Besides giving confidence that the recreated scene accurately represents the real scene, this registration enables the same quantitative evaluations that are shown for the VanLare scene. In this case, the two images are registered to sub-quarter-pixel accuracy. Figure 55 shows the good matches in green and Table 7 shows the resulting transformation matrix along with the sub-quarter-pixel average RMSDE and 295 remaining matches.

Figure 54. Top left: The input FLDOT RGB imagery over Melrose, Florida, Top right: The RGB imagery using the Approach 1 scene, Bottom: The RGB imagery created using the Approach 3 scene.
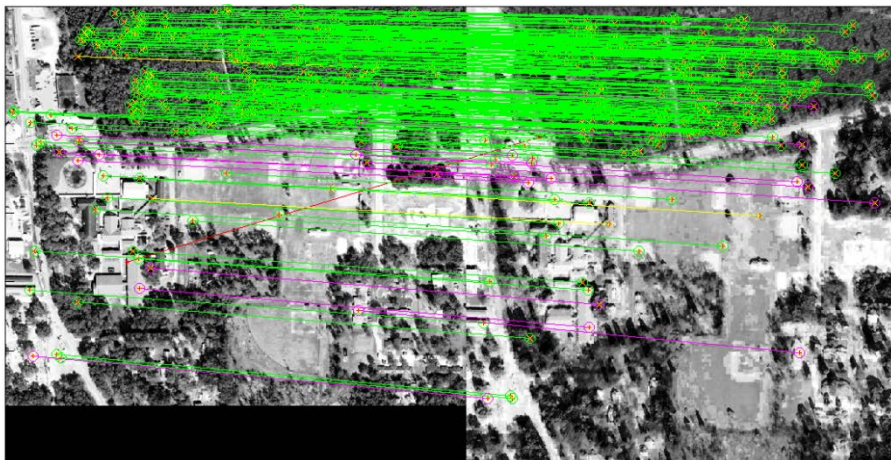
Figure 55. The registration of the input RGB over Melrose, FL to a DIRSIG generated RGB using the recreated scene.

Table 7. The resulting transformation matrix between the input and Approach 1 recreated RGBs over Melrose, FL.

High res RGB Original to DIRSIG Registration		
1.0470	-0.0169	0
0.0066	0.9072	0
6.7381	-5.6218	1
RMSDE Average: 0.2457 pixels		
Matches Remaining: 295		

After registration, the synthetic imagery from the Approach 1 and Approach 3 scenes are compared to the input images in the same way as in the pure modeled VanLare scene. Figure 56 shows the intensity difference per band between the DIRSIG generated RGB and the input FLDOT RGB using the Approach 1 scene and the Approach 3 scene.
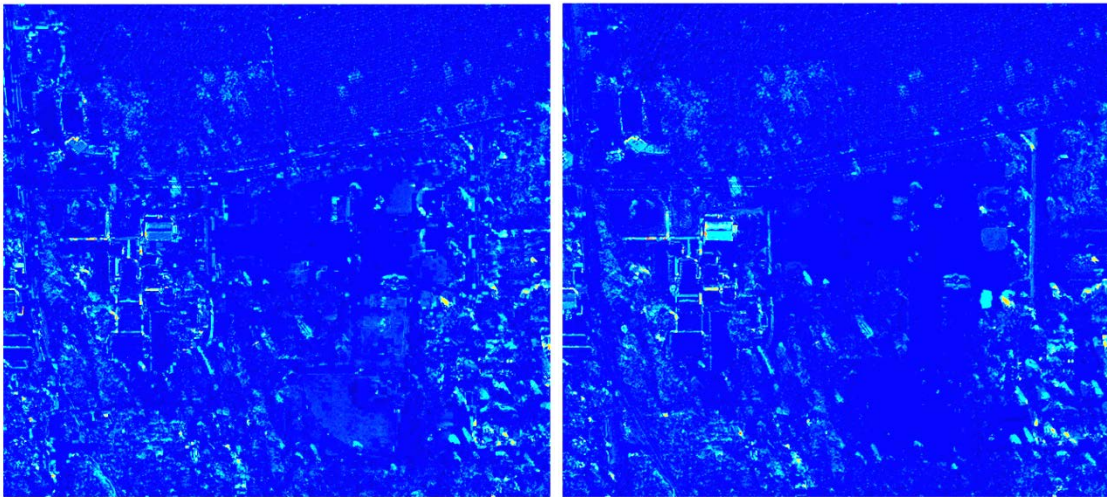
Figure 56. Left: the resulting average band difference image comparing the RGB image of the low-resolution Approach 1 scene to the input FLDOT RGB, Right: the resulting average band difference image comparing the RGB image of the HR Approach 3 scene to the input FLDOT RGB.

In this case the low-resolution Approach 1 scene results in an average of 17.7 digital counts difference per band while the HR Approach 3 scene results in an average of 16.8 digital counts difference per band from the original FLDOT RGB image on the 0 to 255 digital count range. In both the low-resolution and HR scenes, the larger intensity

errors again occur primarily along sharp elevation areas and where trees are not properly separated from ground and building returns in the geometry and thus create hard shadows where there are none in the original imagery.

As in the VanLare case, the DIRSIG generated HSI for the low-resolution and HR scenes can be compared to the input AVIRIS HSI using the average band difference and spectral angle metrics. Figure 57 shows the calculated average reflectance difference per band for the low-resolution Approach 1 scene HSI (top left) and HR Approach 3 scene HSI (top right) as well as the SAM images for the low-resolution scene HSI (bottom left) and HR scene HSI (bottom right). The input HSI is in reflectance so each band has a range of 0 to 100 percent reflectance and the image mean for the average reflectance difference per band for the low-resolution scene HSI is 7.7% while the image mean for the HR scene is 7.9%. The spectral angle image mean for the low-resolution scene is 0.17 radians and 0.18 radians for the HR scene. While the numbers do not indicate a large difference in the accuracy of the imagery resulting from the two scenes, the relatively low numbers do show that both scenes are good representations of the original scene. As in the fully-modeled VanLare dataset case, the general results again show better spectral agreement when intensity is not included particularly away from material geometry transitions in the scene.
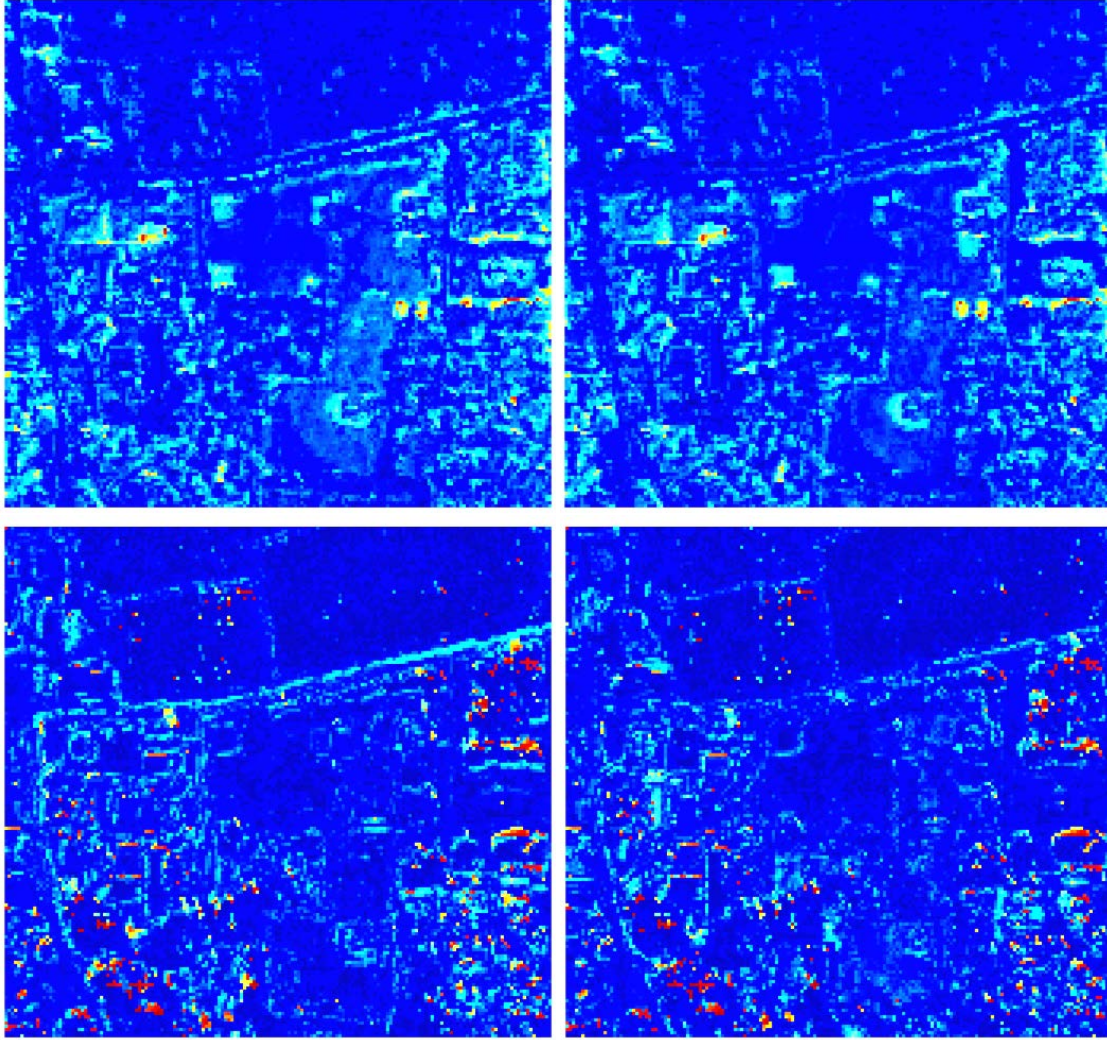
Figure 57. Top left: The average reflectance difference per band for the low-resolution scene HSI, Top right: The average reflectance difference per band for the HR scene HSI, Bottom left: The SAM image for the low-resolution scene HSI, Bottom right: The SAM image for the HR scene HSI.

The consolidated results for Approach 1 and Approach 3 are shown in Table 8 for the VanLare scene and Table 9 for the Melrose scene. The results for both the non-fused, low-resolution (Approach 1) and fused, HR (Approach 3) approaches are relatively low and thus both methods appear to be viable for accurately reproducing the scenes. Determining which approach is better is therefore left to the qualitative evaluation and

104

user requirements. However, if the user intends to model a sensor with higher resolution than the input HSI, Approach 3 does qualitatively give more accurate results.

Table 8. The consolidated results for the two fully automated LHD approaches (Approach 1 and Approach 3) on the VanLare dataset.

Approach	RGB ABD (Digital Counts)		HSI ABD (Percent Reflectance)		HSI SAM (Radians)	
Statistic	Avg	Std	Avg	Std	Avg	Std
1: HSI-only	4.5	6.5	2.0	2.0	0.05	0.07
3: Segmentation	4.5	6.2	2.1	2.3	0.05	0.07

Table 9. The consolidated results for the two fully automated LHD approaches (Approach 1 and Approach 3) on the Melrose dataset.

Approach	RGB ABD (Digital Counts)		HSI ABD (Percent Reflectance)		HSI SAM (Radians)	
Statistic	Avg	Std	Avg	Std	Avg	Std
1: HSI-only	18.3	14.0	7.7	5.2	0.2	0.2
3: Segmentation	17.0	13.5	7.9	5.3	0.2	0.2

5.4 High-Resolution Results and Discussion of the ABD Metric

The inability of the ABD metric to distinguish between the quality of scenes produced using the fused Approach 3 versus the non-fused Approach 1 seems to be in contradiction with evident pixelation error in synthetic imagery produced from Approach 1 scenes. This pixelation error is particularly evident when small GSD synthetic imagery is created using the Approach 1 scenes as shown in Figure 58. This discrepancy led to the investigation of an additional metric in which the number of pixels with error below a set threshold are computed for each of the two scenes. In this metric, the ABD value at each pixel is divided by the illumination intensity of the input imagery at that pixel where the grayscale illumination intensity is found using MATLAB's rgb2gray function (MATLAB, 2010). The resulting image is a percent error image which is shown for a windowed region of the VanLare scene in Figure 59 for Approach 1and Approach 3. Using this metric on the windowed region, the Approach 1 windowed region has 70.4%

105

of its pixels within 15% of the original input RGB image while the Approach 3

windowed region has 86.5% of its pixels within 15% of the original input RGB image.

Over the same windowed region, the ABD values are still relatively similar at 6.7 digital

counts for Approach 1 and 5.8 digital counts for Approach 3.

From this new thresholded percent error metric, it can be concluded that the

sharpening of the material map accomplished by Approach 3 pushes large errors into

small regions. Thus, it does create a higher resolution and more accurate material map,

but similar amounts of overall error are redistributed to smaller boarder regions between

materials. The windowed and thresholded percent error metric provides an additional

quantitative evaluation of the recreated scenes and could be further pursued in future

research to help distinguish between automated synthetic scene generation approaches.

5.5   Chapter Summary

The qualitative evaluation of synthetic imagery generated from the recreated

scenes of Approach 1 and Approach 3 along with registration of the synthetic imagery to

the input imageries lend high confidence to the automated LHD method. Furthermore,

while determining a good metric to quantify the absolute accuracy of a reconstructed

synthetic scene is a difficult task, the average band difference and spectral angle metrics

at least allow for the ability to quantify relative accuracy within and between scenes. The

relatively low numbers given by these metrics when applied to the VanLare and Melrose

scenes recreated with both the low-resolution non-fused material map of Approach 1 and

the HR fused material map or Approach 3 give a high confidence that the automated

extraction of geometry, texture and spectra is completed with a sufficient level of

accuracy using the LHD method. The initial results of the windowed and thresholded

percent error metric are promising, but additional research is needed to improve the

quantitative evaluation of autonomously generated synthetic scenes.

Figure 58. Small GSD imagery of the Approach 1 (top) and Approach 3 (bottom) generated scenes.
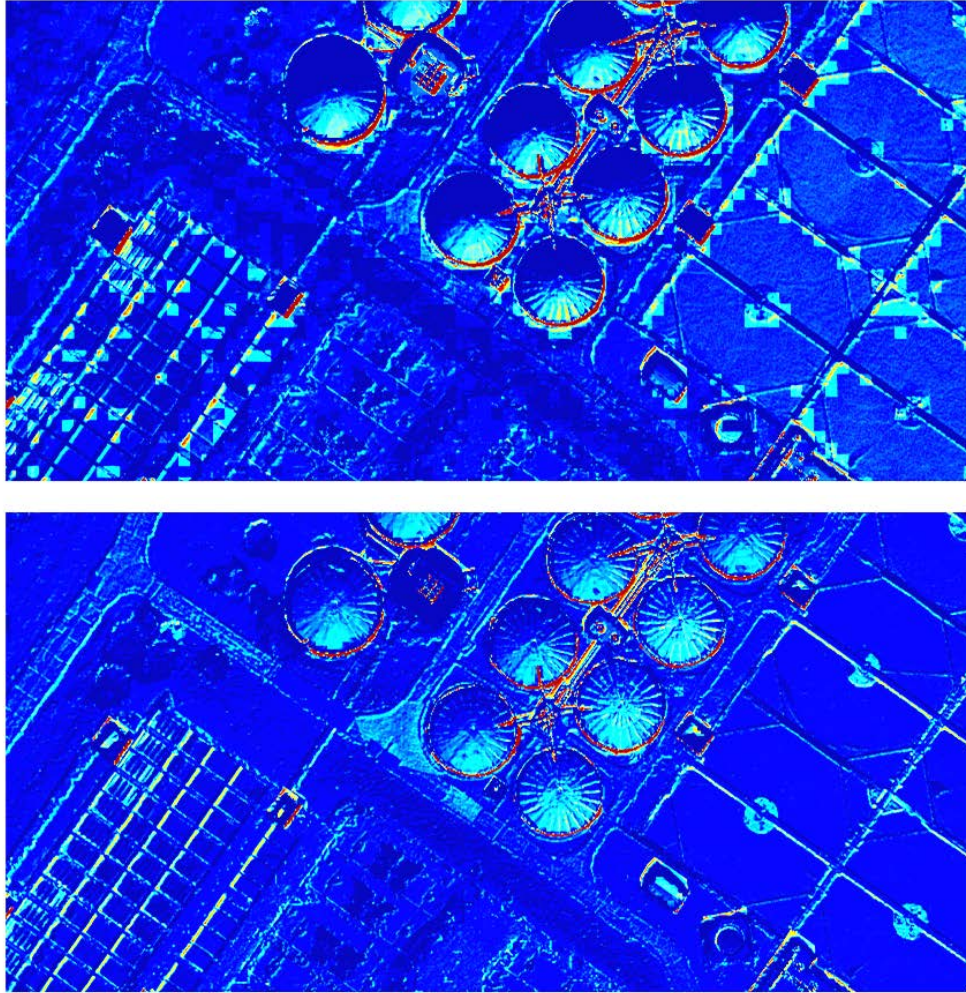
Figure 59. Percent error image for the Approach 1 scene (top) versus the Approach 3 scene (bottom) for a windowed portion of the VanLare dataset.

VI. Conclusions and Recommendations

This chapter discusses the implications of the methods developed in this research and their results in the modeling community. It begins by identifying and summarizing the specific contributions made in the research. Next, it presents some of the future research paths which could be taken to further advance the research. The final section closes with a summary of the chapter and overall effort.

6.1   Research Contributions

Fully automating the generation of synthetic scenes is the initial and ultimate goal of this research, but the automated processes which are presented here to accomplish automated synthetic scene generation are not the only resulting contributions. The following sections present the specific contributions resulting from this research effort.

6.1.1   Full Automation of Synthetic Scene Generation

Previous research efforts greatly reduce the time required to generate synthetic scenes by automating the extraction of the scene geometry and texture from lidar and HR imagery. Generating wide-area scenes using the methods developed in those efforts can be accomplished in days instead of months. The assignment of spectral content to the scenes, however, is still a manual effort when using these previous methods. Furthermore, when using a manual method to assign spectral content, if assigned spectra are extracted from a material library, the user may be limited to the assignment of a best-match representative spectrum of a similar material rather than the exact material to be modeled. Thus, the assigned spectra may not be perfect representations for the materials

in the scene and may not include the true spectral diversity of the materials at the real-world site.

The LHD method developed in this research automates the spectral assignment in the scene reducing the time required to generate a scene from days to hours and using scene-measured spectra pulled directly from the HSI. Additionally, since no commercial products are used in the geometry extraction developed in this research, the algorithms are written in MATLAB and are freely available. Finally, because the resulting material map can be related to the texture maps using the transformation matrix from the registration step, individual mean and standard deviations can be calculated for each material in the scene. This allows DIRSIG to better select spectra from the material library which best match the red, green, and blue intensity values of the texture maps. The bullets below highlight some of the main contributions of the LHD method:

- Full automation reduces time required to generate a scene and allows construction of a scene even if it cannot be accessed by ground
- Deriving the spectral library directly from HSI ensures the scene spectra match the materials of the site and capture real-world variety and complexity
- Geometry extraction algorithm is written in MATLAB and available for future research (no additional commercial software required)
- The constructed scene allows for registration of disparate modalities and look angles to be registered to the scene accounting for three-dimensional effects acting as a Rosetta Stone for the site
- Material map is used to pull material specific intensity means and standard deviations from the texture maps enabling more accurate spectral assignment

6.1.2   Scene Classification: New Fused Approaches

Besides contributing to the synthetic scene, synthetic imagery, and physics-based modeling community, the three approaches developed to automate the spectral assignment step of the LHD method provide new contributions to the general field of

scene classification. With multi-modal collects becoming more common, interest in data fusion and fused classification is increasing. Besides gaining access to new wide-area scenes for algorithm testing and development, the primary interest of the sponsor of this work (AFRL/RYA) is the ability to fuse and extract information from multiple modalities. All three approaches used to incorporate spectral information into the scene generation process take advantage of Walli's SIFT, RANSAC, average RMSDE minimization registration process and show that this robust registration process can be used to relate the information of the HR imagery, lidar, and HSI. Approaches 2 and 3 leverage the resulting registration transformation matrices to improve the resolution of the scene classification with novel fused approaches. The bullets below highlight the main contributions in this area of the research:

- HSI can be registered to rasterized lidar return strength images and HR imagery using the Walli registration process
- Using the rasterized lidar elevation image as a band in a fused cube allows non-correlated information to be added to the classification process
- The fused RGB and lidar cube enables the creation of material-specific, five-component description vectors
- By segmenting the scene at high resolution then incorporating HSI information, Approach 3 shows an approximately 10-fold resolution enhancement in the resulting material map

6.1.3   Evaluation of Scene Accuracy

Another area of contribution for this research effort is in the ability to quantitatively evaluate the accuracy of the autonomously generated scene by comparing scene-generated outputs to the scene inputs. While the ABD metric did not show a significant difference between the Approach 1 and Approach 3 results, it does give an

111

intuitive metric to understand the accuracy of the output synthetic imagery even in the multi-dimensional space of the RGB imagery and HSI.

The additional proposed windowed and thresholded percent error metric provides a good first step and quantitatively shows a difference between the Approach 1 and Approach 3 results, but still relies on a somewhat arbitrary accuracy threshold. It gives the community an additional starting point to generate better metrics for the comparison of output to input imagery for the evaluation of synthetic scene accuracy regardless of the method used to generate the scene, but falls short of providing a full evaluation of the scene's ability to produce synthetic imagery which accurately represents real-world data.

6.1.4 Suite of MATLAB Software Tools

A DVD with the documented MATLAB code to read in the input files, extract the information, and output the DIRSIG scene files will be submitted to committee members and the AFIT Engineering Physics Department. Also included will be readme instruction files and the scene inputs and outputs. These DVDs will be provided to each of the committee members and the AFIT Engineering Physics Department for use by future students or interested researchers. Some of the highlights of the included imagery and code are listed here:

- Input RGB imagery, lidar data, and HSI for both the modeled VanLare scene and real Melrose, FL scene
- Walli's two-dimensional registration toolkit incorporating epipolar and homographic constraints, and average RMSDE minimization
- Geometry extraction toolkit to read in the lidar, separate tree canopies, and output .obj files
- The SEM (modified from Dr. Eismann's original code) and ISODATA algorithms used for scene classification
- MATLAB scripts to walk through the additional fused scene classification steps

112

6.2 Recommendations for Future Research

Limitations of the models developed using the LHD method are readily apparent under a few important circumstances. First, when sensors with GSDs smaller than one meter are used to image the scene, the relatively crude facetization of the scene geometry is visible. Refinement techniques developed by Lach and others would help to reduce the number of facets used for man-made structures as well as improve tree representations. Also, texture and spectral information is not available for vertical facets with the LHD approach because it does not yet incorporate off-nadir inputs. Thus the scene is limited to modeling near nadir geometries only. Finally, any shadowing in the texture map leads to the selection of darker representative spectra for the materials that fall in the shadow, so shadows from the texture map show up in the final model regardless of the sun angle being modeled by the user. Thus, the scene exhibits multiple shadows if the scene is used to model sun angles which vary significantly from the sun angle captured in the texture map. Each of these limitations are apparent in the close-up RGB image of the elementary school in the Melrose scene shown in Figure 60.

Thus, future work should focus on reducing these limitations by improving scene geometries using available refinement techniques. A method could be developed in which tree type, location, and size could be determined from the input spectral and elevation information. This information could then allow the planting of an appropriate modeled tree to fit those parameters. The shadowing of the texture maps could be addressed by using the actively illuminated, rasterized lidar return strength as the texture map, although this would reduce the GSD of the texture map.

Figure 60. Close-up off-nadir RGB image produced using the HR Approach 3 recreated Melrose scene.

Incorporating off-nadir imagery to supply texturing of vertical facets would also be an interesting challenge and would extend the scene's utility beyond near-nadir modeling. Additionally, investigating sensor pointing error and noise effects on the LHD method would provide good insight into how robust the method is for other various input sources. Finally, linking HSI class spectra to a lab measured spectral library would allow for the improvement of reflectance spectra in the scene as well as the ability to extend the spectral range of the scene past that of the input HSI.

Even though the linear unmixing method used in Approach 2 to generating a HR material map showed limited utility with the two datasets used in this research, the intuitiveness and physicality of this approach are appealing. Additional research could be

applied to this method both in determining a reliable automated method to select endmembers and in determining the subpixel endmember locations within the HSI pixels using the HR information from the rasterized lidar and RGB imagery. Additionally, it may be that the two datasets used for this research are coincidentally better suited to a normal mixture model clustering approach rather than the linear mixture model endmember approach. Additional datasets could be used to determine the utility of Approach 2.

The average band difference metric provides an intuitive evaluation of the scene-wide error and the windowed and thresholded percent difference error quantifies the sharpening improvement of Approach 3 versus Approach 1, but neither are a perfect metric for the evaluation of the utility of a synthetic scene. The true value of a synthetic scene lies in its ability to be used in a physics-based model to produce synthetic imagery which accurately predicts real imagery of the same or a similar site. For example, if the scene is to be used to evaluate the capability of target detection algorithms to find targets in a scene, the performance of those algorithms when used on the synthetic imagery should be equivalent to the performance of the algorithms when the real sensor captures imagery over the same or similar sites. Alternatively, if a scene is constructed to serve as a Rosetta Stone, then it should have the ability to produce synthetic imagery good enough to be registered to imagery not used in the construction of the scene. Unfortunately, additional modalities for the Melrose scene are not currently available for this test. Thus additional research is needed in this area to help quantify the real utility of synthetic scenes whether they are generated manually or autonomously.

6.3 Summary

This research presents a fully automated method to extract the three necessary components of a wide-area synthetic scene: Geometry, texture, and spectral. Whether using the non-fused Approach 1 or fused Approaches 3 to generate the material map, the synthetic images produced using the LHD-generated scenes are realistic both qualitatively and quantitatively lending confidence to this new method. Additionally, the automated incorporation of HSI allows the extraction of high spectral resolution information expanding beyond the visible wavelengths and does so accurately and more quickly than previous manual and semi-automated methods. This advancement provides the synthetic imagery modeling community with a new, faster method for scene generation even when ground access to a site is not available and also provides the potential to create a new library of scenes featuring a wider array of various terrain and vegetation than currently exists.

Appendix A – Determination of Texture Map and Material Map Insertion Point

Relating the various coordinate systems of all of the input data to the final coordinate system in DIRSIG requires some careful consideration. As is shown in Appendix 2, one of the scene inputs for DIRSIG is the latitude and longitude for the origin of the coordinate system of the scene. This is the lower left-most vertex of the lidar-based facetized geometry. Thus the minimum values for the input x and y lidar coordinates over the region of interest, which were in Universal Transverse Mercator (UTM) coordinates in this case, are subtracted from the lidar dataset through an affine translation as shown in the Melrose example in Figure 61 to covert from the lidar coordinate system to a local DIRSIG coordinate system.

UTM Coordinates, Zone 17   Lat, Long: (29.70°, -82.05°)

$$[398779, 3286479, 1] * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -398779 & -3286479 & 1 \end{bmatrix} = [0, 0, 1]$$
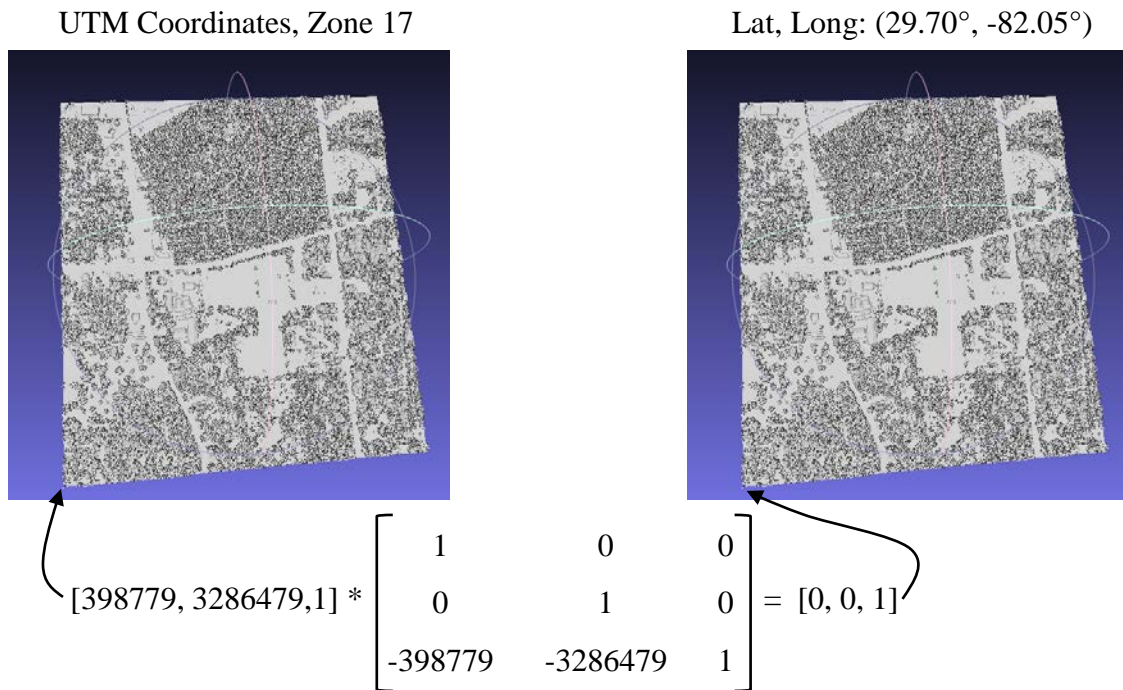
Figure 61. Applying a translation to shift the origin on the Melrose scene geometry from UTM coordinates to a local DIRSIG coordinate system.

117

The x and y lidar coordinates in UTM are in units of meters from the origin of zone 17. The latitude and longitude coordinates of the origin were found using a conversion website provided at http://www.rcn.montana.edu/Resources/Converter.aspx then entered into the DIRSIG scene generator Graphical User Interface (GUI).

With the local coordinate system of the scene geometry set up in DIRSIG, all that remains is to determine the placement of the texture and material maps relative to the geometry for proper draping of these maps over the geometry. This involves the careful tracking of all of the affine transformations involved in the image preparations for registration as well as the determined registration matrices to come up with the composite transformation matrix to convert the texture and material map origins to the local DIRSIG coordinate system. When performing using the HR method of LHD, the texture map and the material map are in the same coordinate system, so they have the same insertion point into DIRSIG. This origin is found by transforming the origin on the original HR image (0,0) through all of the transformations to arrive at the location of the transformed origin in the local DIRSIG coordinates. The equation for this process is

$$[0,0,1] \times \begin{bmatrix} High\ Res \\ to\ HSI \\ GSD\ Ratio \end{bmatrix} \times \begin{bmatrix} High\ Res \\ to\ HSI \\ Registration \\ Transform \end{bmatrix} \times \begin{bmatrix} Lidar\ to \\ HSI \\ Registration \\ Transformation \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} Lidar\ to\ HSI \\ GSD\ Ratio \end{bmatrix}^{-1} \times \begin{bmatrix} Lidar\ Translation \\ to\ Upper\ Left\ Most \\ UTM\ Coordinate \end{bmatrix}^{-1} \times \begin{bmatrix} Flip\ of\ Y\ axis \\ for\ Cartesion \\ to\ Image \\ Coordinates \end{bmatrix}, \qquad \text{(A1.1)}$$

where the lidar translation comes out to zero for the x coordinate, since the lidar has already been translated to have a (0,0) origin in the DIRSIG coordinate system, and the maximum y value to get to the upper-left-most geometry vertex for the y coordinate since

the texture and material maps were placed using an image coordinate system. It should

be noted that two different origins are in use here. The geometry was placed using a

Cartesian coordinate system and thus has its origin at the lower-left-most vertex and

positive x increases to the right and positive y increases in the up direction. The texture

map and material maps were placed using an image coordinate system where the origin is

located in the upper left and positive x increases to the right and positive y increases in

the down direction. Although not done here, in hindsight, it would have simplified this

process of determining the origin by using the DIRSIG option of placing the texture and

material map using a Cartesian coordinate system.

   One final consideration should be noted that in the case where either the texture

map or material map have a non-zero rotation when registered with the lidar, the original

origin of the draped maps does not give the correct origin in the local DIRSIG space

because of image padding. If a counter-clockwise rotation was needed in the registration,

the x-value is given by the upper-left point, but the y value comes from the upper-right

point. For a clockwise rotation, the x value of the new origin comes from the lower-left

point and the y-value from the upper-left. Figure 62 shows this process pictorially for the

counter-clockwise case (top) and clockwise case (bottom).
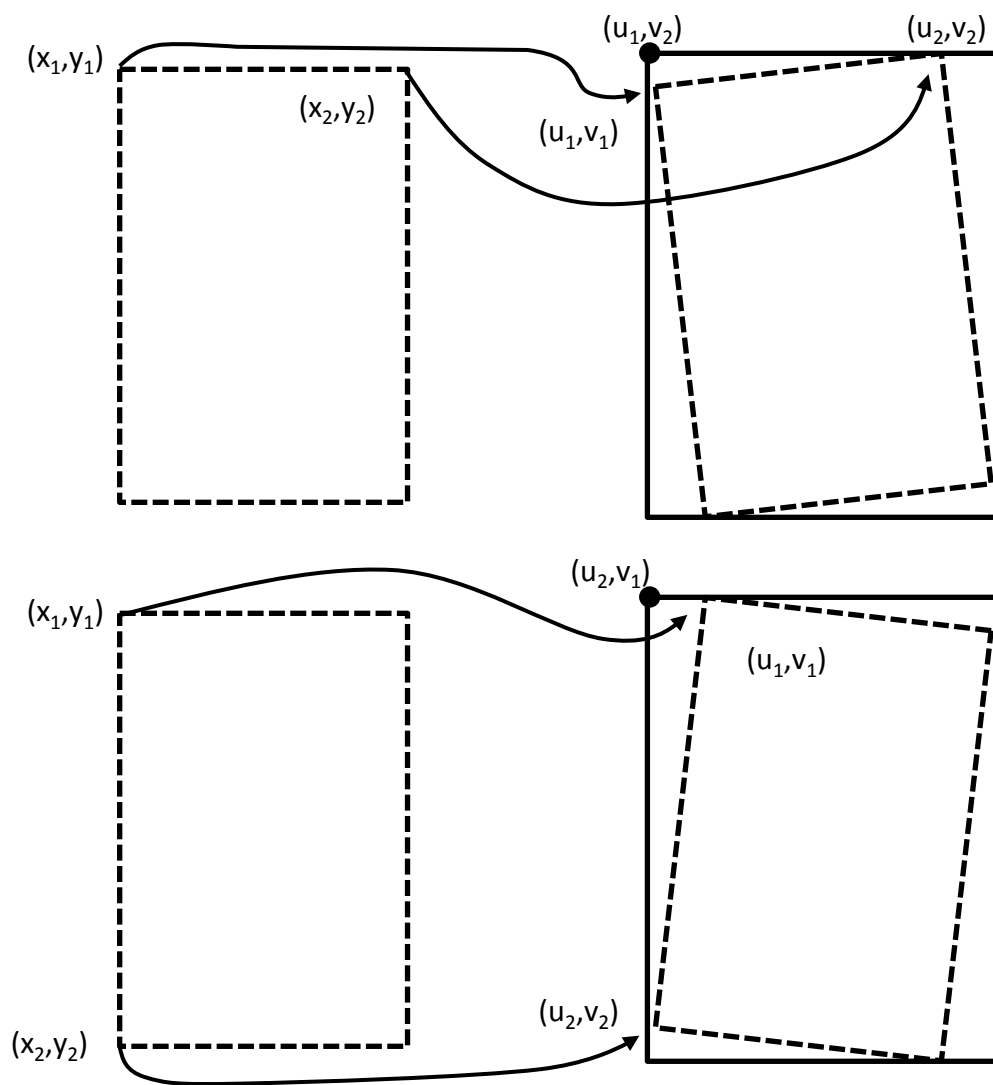
Figure 62. Determining the upper left insertion point when a counter-clockwise (top) or clockwise (bottom) rotation is involved.

Appendix B –Building a Scene Using the DIRSIG GUI

While the LHD method automates the extraction of the scene components in the formats necessary for DIRSIG, the DIRSIG Graphical User Interface (GUI) was used to construct the .scene text file which tells DIRSIG where all of the scene component files are located and how the texture and material maps are related to the geometry and material library. The .scene text file could be generated using a MATLAB script, but the GUI is a quick and easy way to accomplish the task. The Melrose scene generated using the HRLHD method is used as an example for the following figures. The first tab in the DIRSIG scene generator is the "General" tab shown in Figure 63.
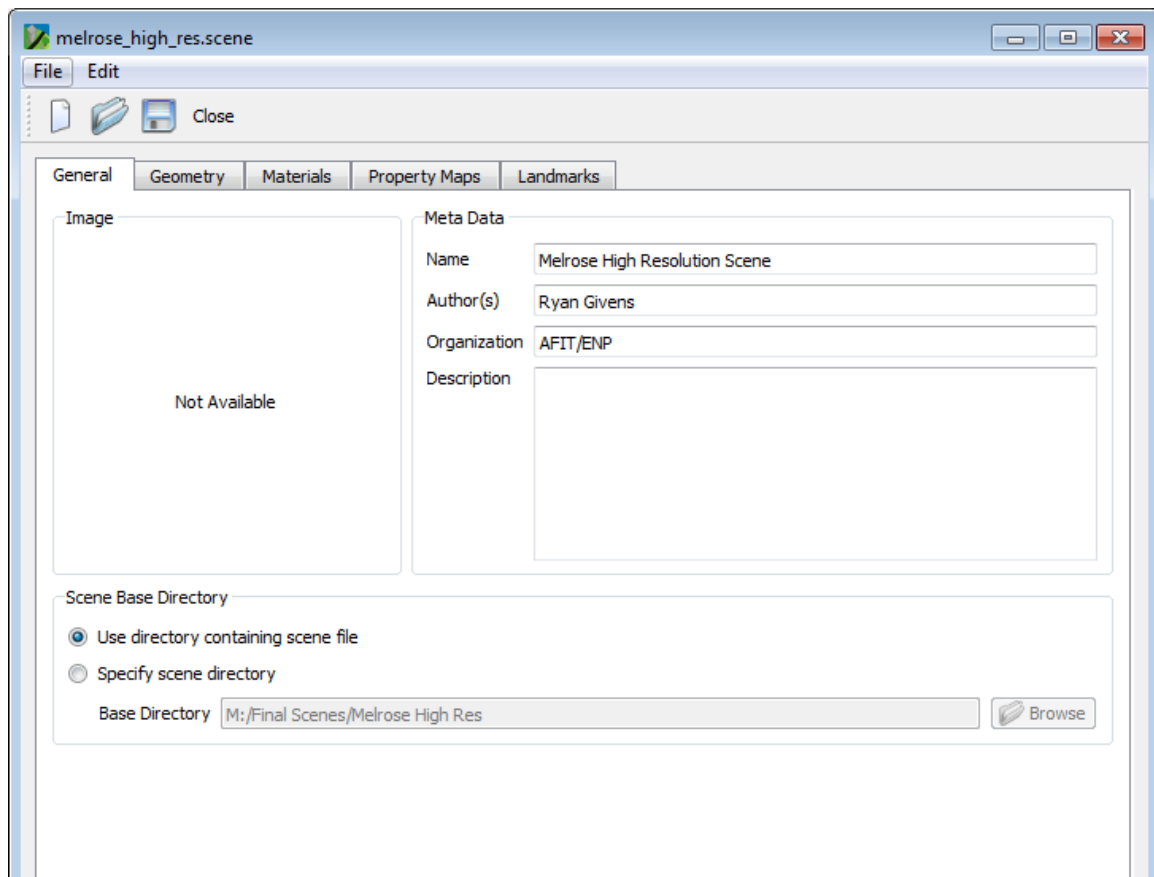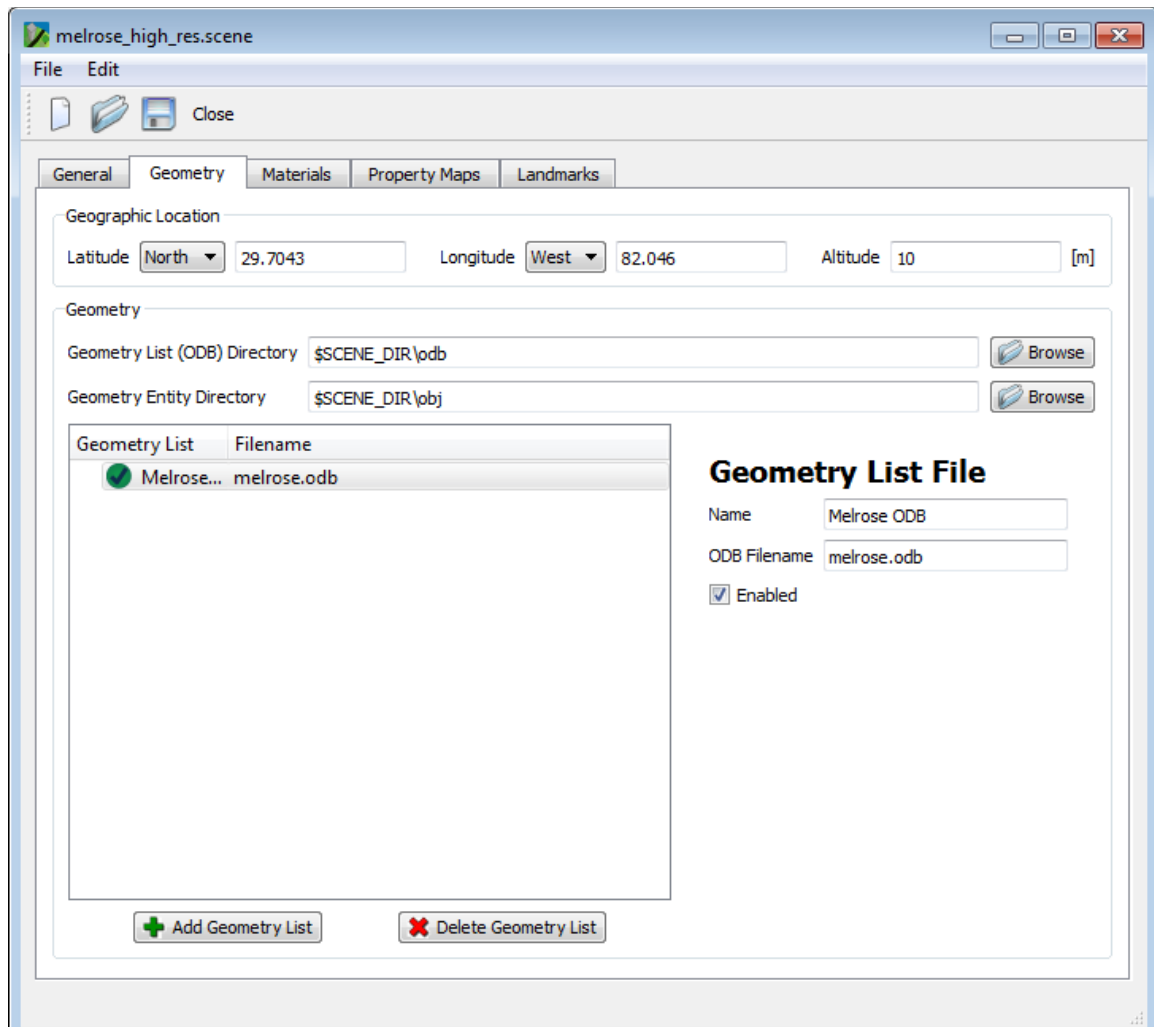

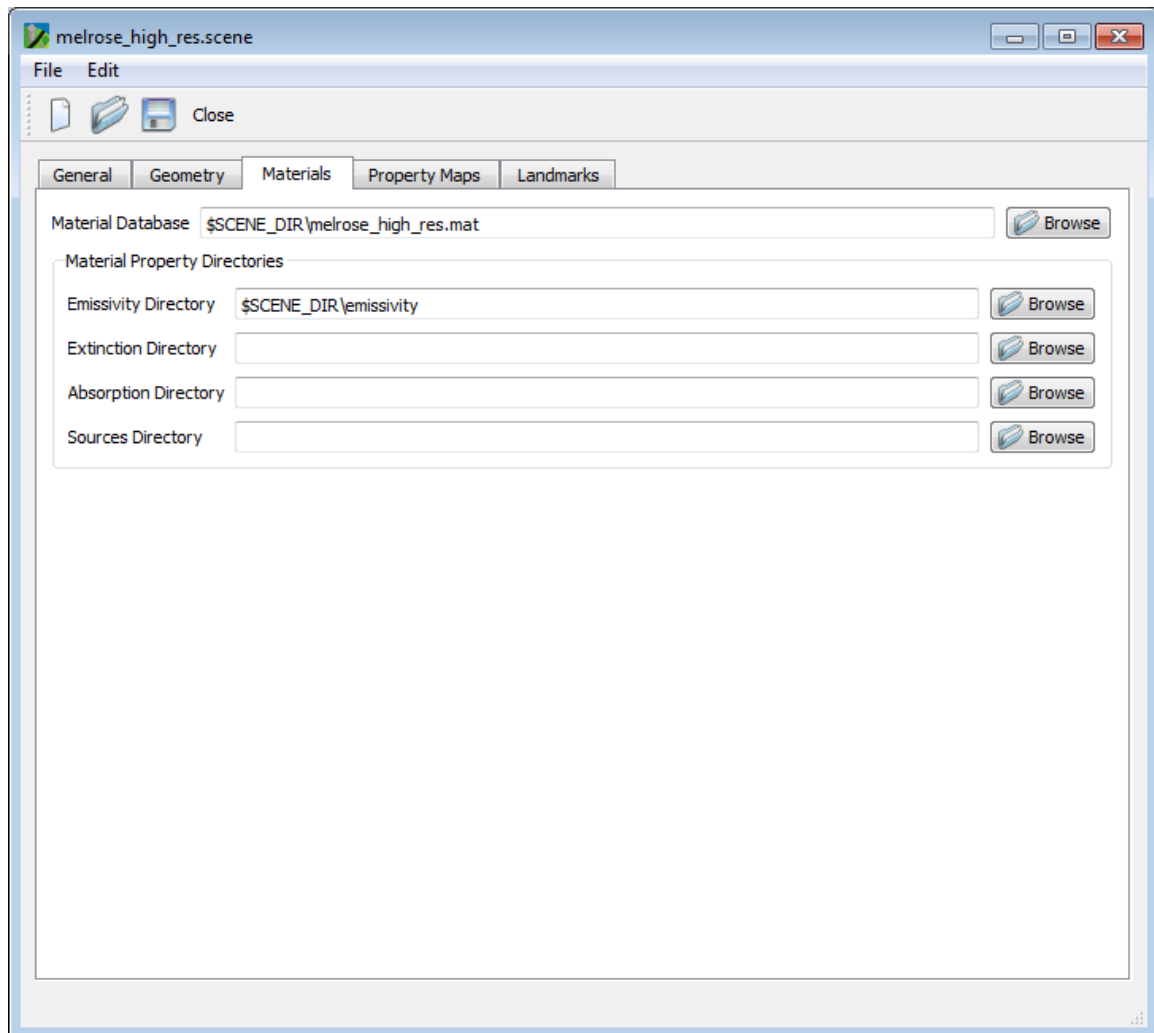**Figure 63. The General tab for the DIRSIG scene GUI.**

The General tab simply allows the user to input the scene name, author and description if desired. The most important option on this tab is the specification of the scene base directory. If portability to another computer is desired, it is best if the user selects "Use directory containing scene file" so that the base directory is not "hard coded" into the .scene file.

The next tab is the "Geometry" tab. This tab allows the user to specify the geographic location of the origin of the local DIRSIG coordinates. This is the latitude and longitude of the lower-left-most vertex of the geometry assuming this point has been shifted to (0,0) using the technique shown in Appendix 1. Next on the "Geometry" tab is the specification of the locations of the Geometry List Directory and Geometry Entity Directory. For the LHD method, the .odb file is a text file specifying the location of the .obj files generated from the lidar input. The .odb file thus contains the file names of the ground and building geometry file and each of the tree return (floating tile) geometry files as well as their insertion point into the scene (which was always at the origin for this case). Also note that the directories were again generalized here for easy portability to other computers by using "$SCENE_DIR\" as the base directory. The screen shot of the "Geometry" tab is shown in Figure 64.

Figure 64. The Geometry tab for the DIRSIG scene GUI.

The "Materials" tab specifies the location of the material database file (.mat

extension) and emissivity directory which is the location of the spectral library. For

scenes constructed using lab measured spectra where surface and bulk properties besides

reflectance are known, these properties are listed in the .mat file. These properties are not

known in the LHD method, however, so the .mat file is simply a list of the materials

pointing to the appropriate emissivity files in the spectral library. The example

"Materials" tab is shown in Figure 65.

Figure 65. The Materials tab for the DIRSIG scene GUI.

       The most complicated tab, as far as the LHD method is concerned, is the

"Property Maps" tab. This tab specifies how the material and texture maps should be

applied to the scene geometry. The first information box specifies the maps directory.

Next, the material map is set up under "Property Maps". The "General" tab under the

material map allows the material map to be named and assigned to the geometry. The

assignment is made using a "dummy" material in the .mat file, material 100 in this case.

In the .obj geometry files, all of the facets including ground, buildings, and tree facets,

are labeled as material 100. This tells DIRSIG that the material map is to be draped over

the facetized geometry. Figure 66 shows the "General" tab for the HR material map in the Melrose scene.
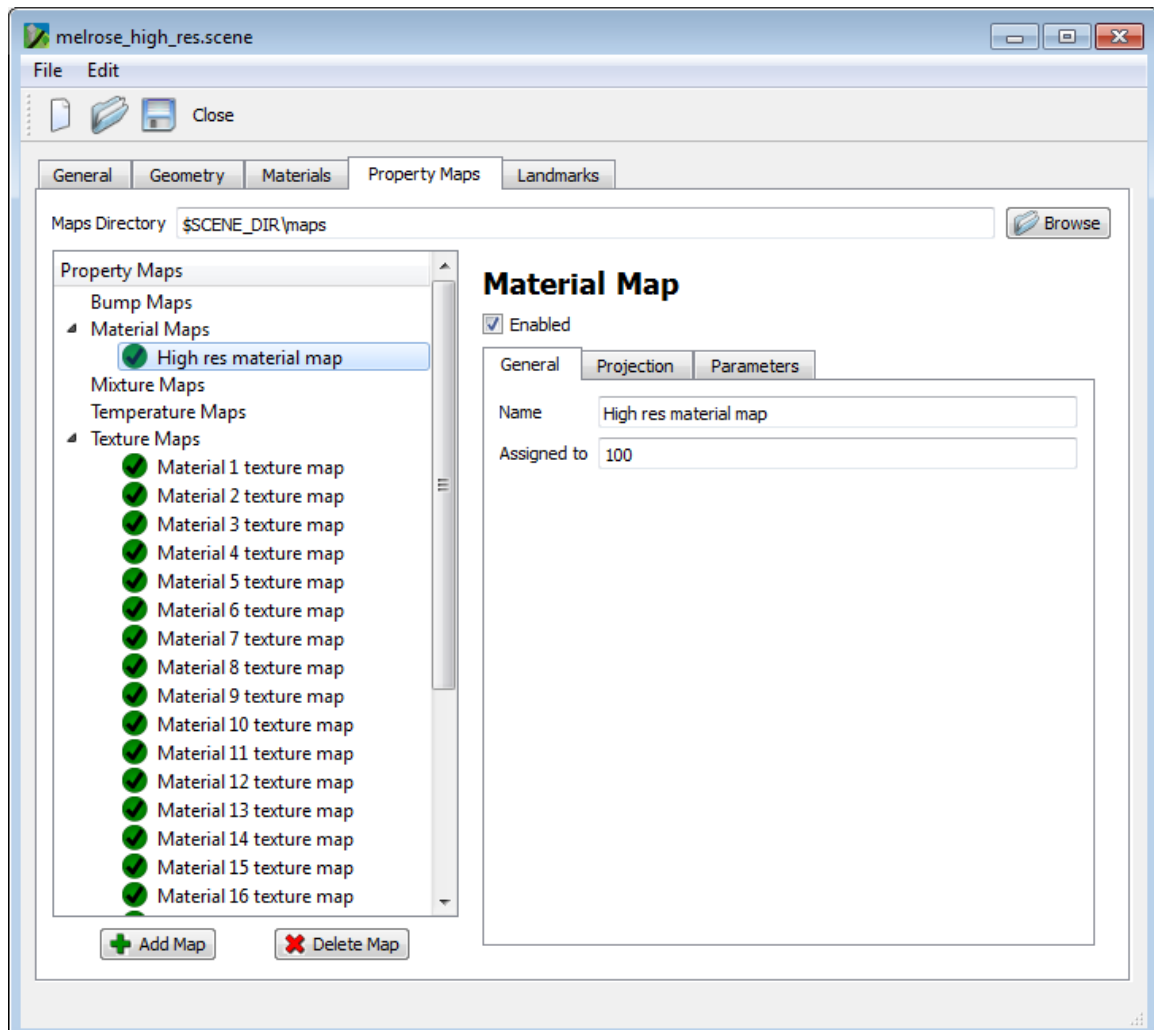
Figure 66. The Property Maps tab for the DIRSIG scene GUI showing the General tab for the material map.

The next tab for the material map is the "Projection" tab. This tab specifies the type of coordinate system to use when draping the material map, how to handle the assignment of materials that fall outside of the material map (with mirroring or repeating of the map), and most importantly for this case, the insertion point and GSD of the

material map as calculated in Appendix 1. This tab is shown in Figure 67 for the Melrose

scene.
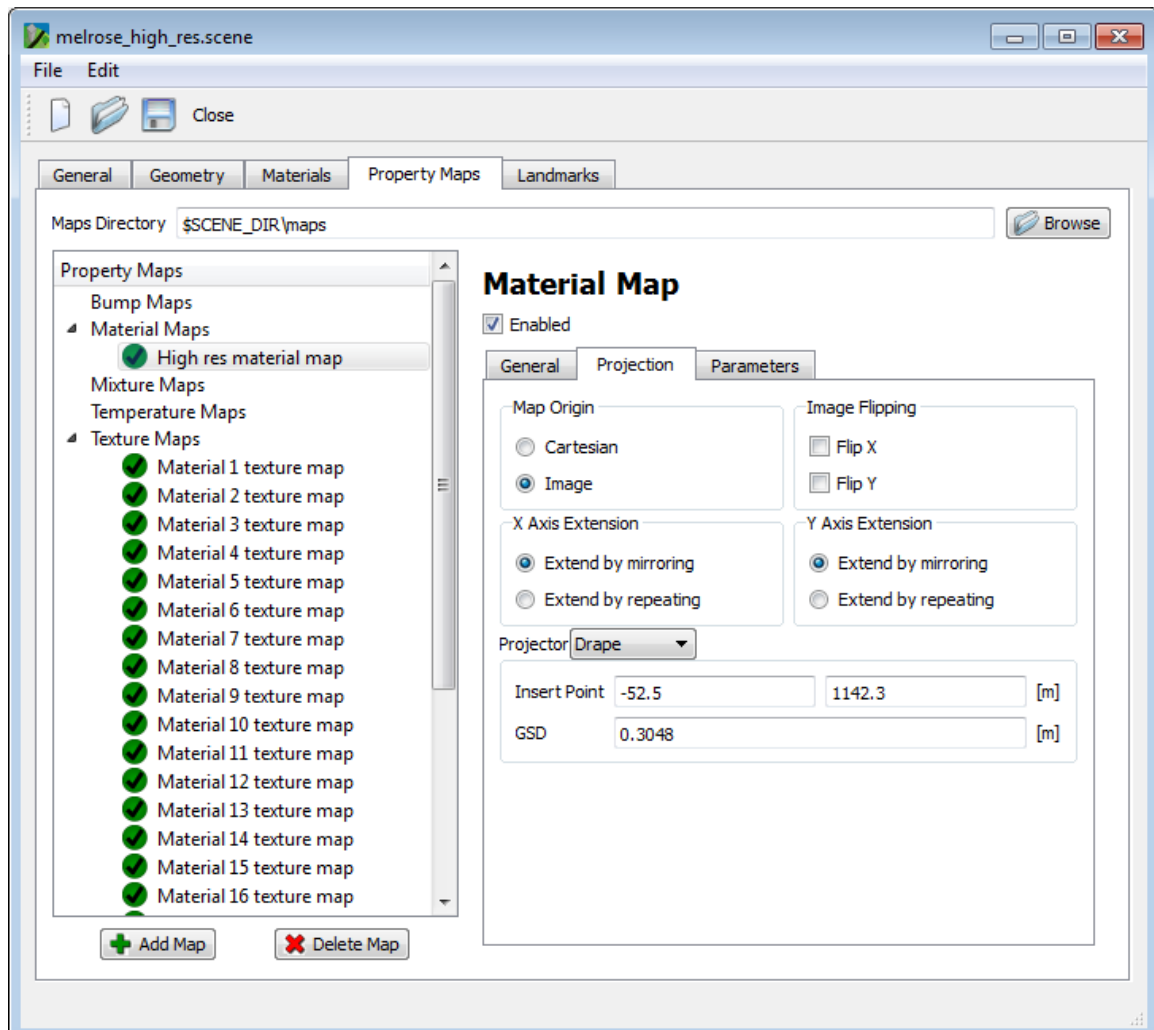
Figure 67. The Property Maps tab for the DIRSIG scene GUI showing the Projection tab for the material map.

The final tab for the material map is the "Parameters" tab. This tab tells DIRSIG

the filename of the material map and how to use the digital counts found in the material

map to assign materials found in the spectral library. Although any numbering scheme

could be used in both the material map and the spectral library, it seemed simplest to

make the digital counts in the material map match the numbering of the materials in the

spectral library. This correspondence is shown in Figure 68.
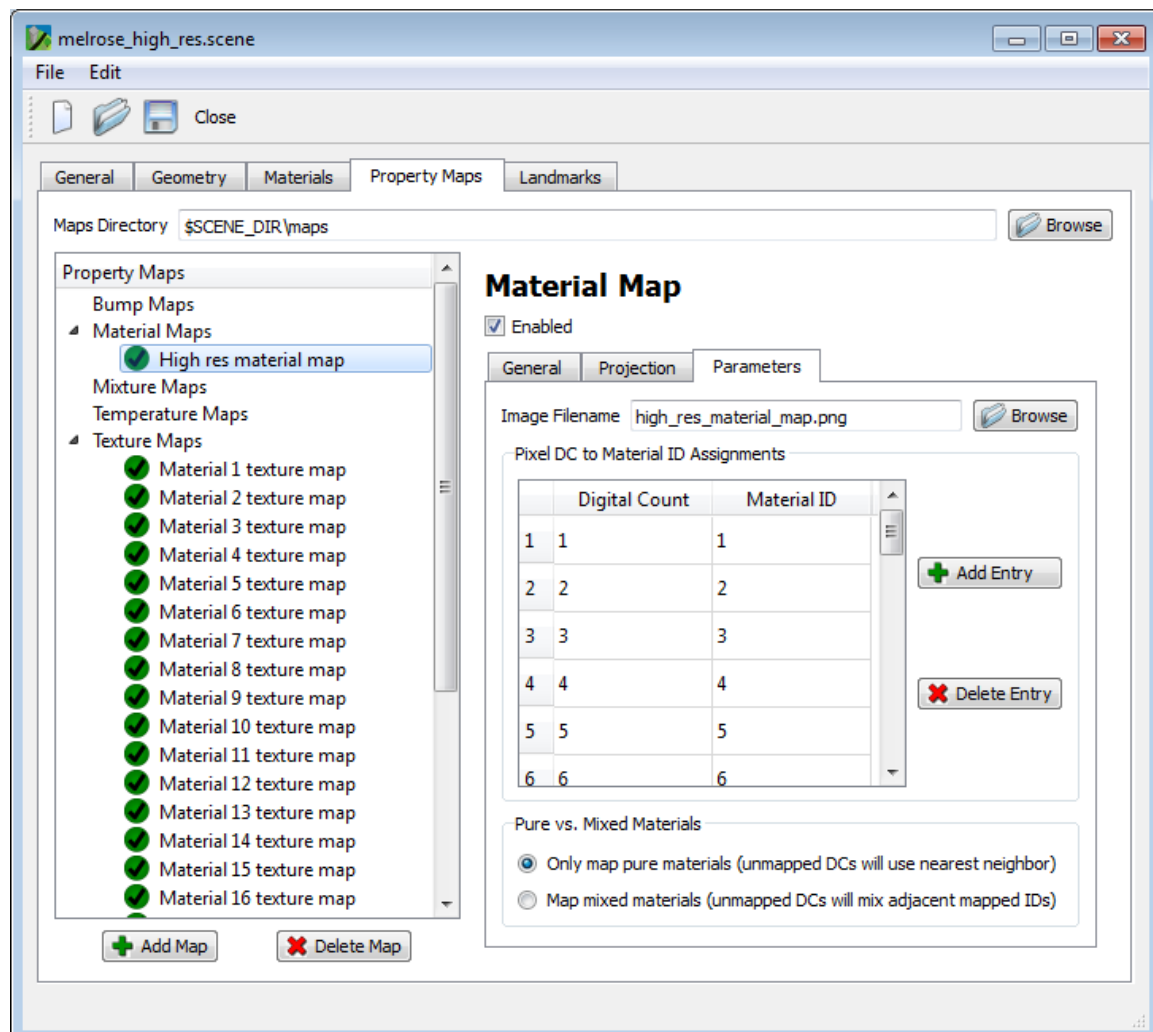
Figure 68. The Property Maps tab for the DIRSIG scene GUI showing the Parameters tab for the material map.

The final step in setting up the .scene file using the DIRSIG GUI is specification

of the texture map properties. Like the material map, the texture map has three tabs:

General, Projection, and Parameters. The "General" tab includes the name of the texture

map and the number of the material to which it is assigned. The example shown in

Figure 69 is the first texture map which was assigned to material 1.
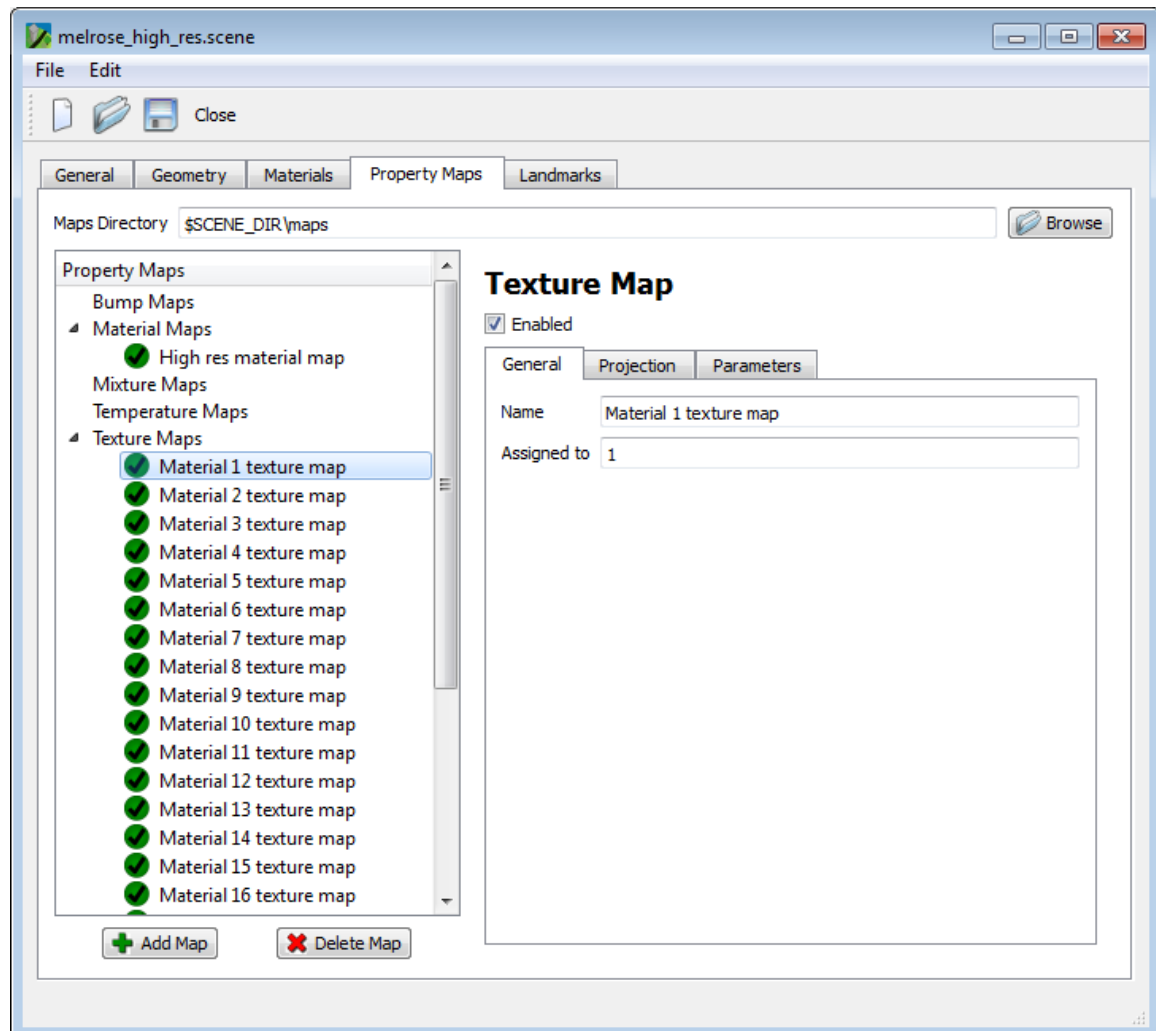
Figure 69. The Property Maps tab for the DIRSIG scene GUI showing the General tab for the first texture map.

In this HR example, since the material map was generated at the resolution of the texture map, the image insertion points and GSDs are equal and the "Projection" tab for the texture map is identical to the "Projection" tab for the material map. Figure 70 shows the example "Projection" tab for the material 1 texture map.
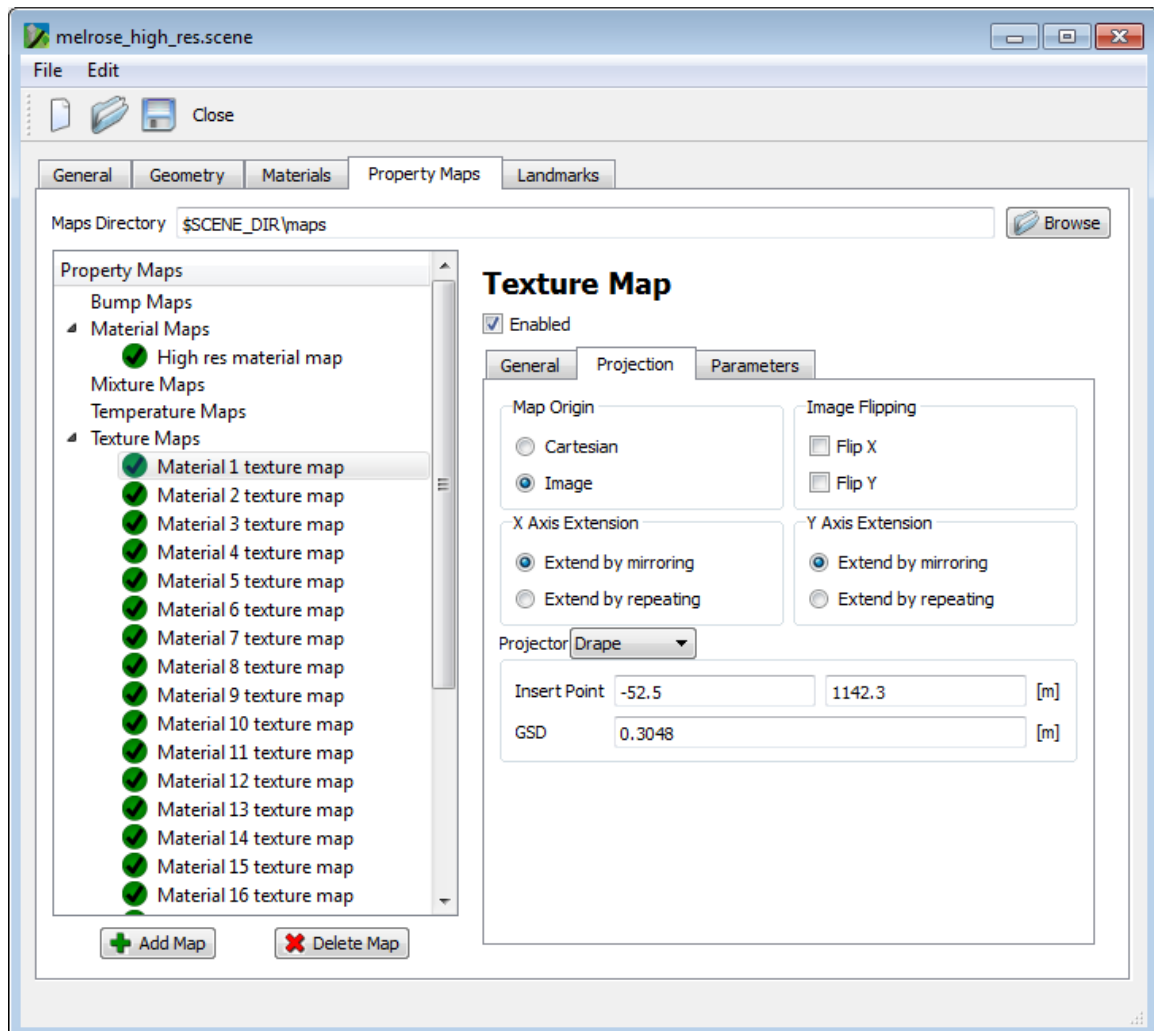
Figure 70. The Property Maps tab for the DIRSIG scene GUI showing the Projection tab for the first texture map.

The "Parameters" tab for the texture map includes the filenames of the texture maps used, the wavelength regions over which they apply, and the material means and standard deviations of the digital counts in the texture maps determined separately for each material in the scene. Thus, while each material in the scene uses the same three texture maps (red, green, and blue), the statistics for each material are different. This was done to ensure that DIRSIG selects the appropriate, and most representative, spectra from each of the spectral library files for each of the materials. If only an image-wide mean

and standard deviation were used and applied to all classes, DIRSIG would select only

the dimmest spectra from materials which were brighter than the rest of the scene (like

white roof tops) and only the brightest spectra from materials which were darker than the

rest of the scene (like black asphalt or dark vegetation). The material specific statistics

for the first material in the HR Melrose scene are shown in Figure 71.
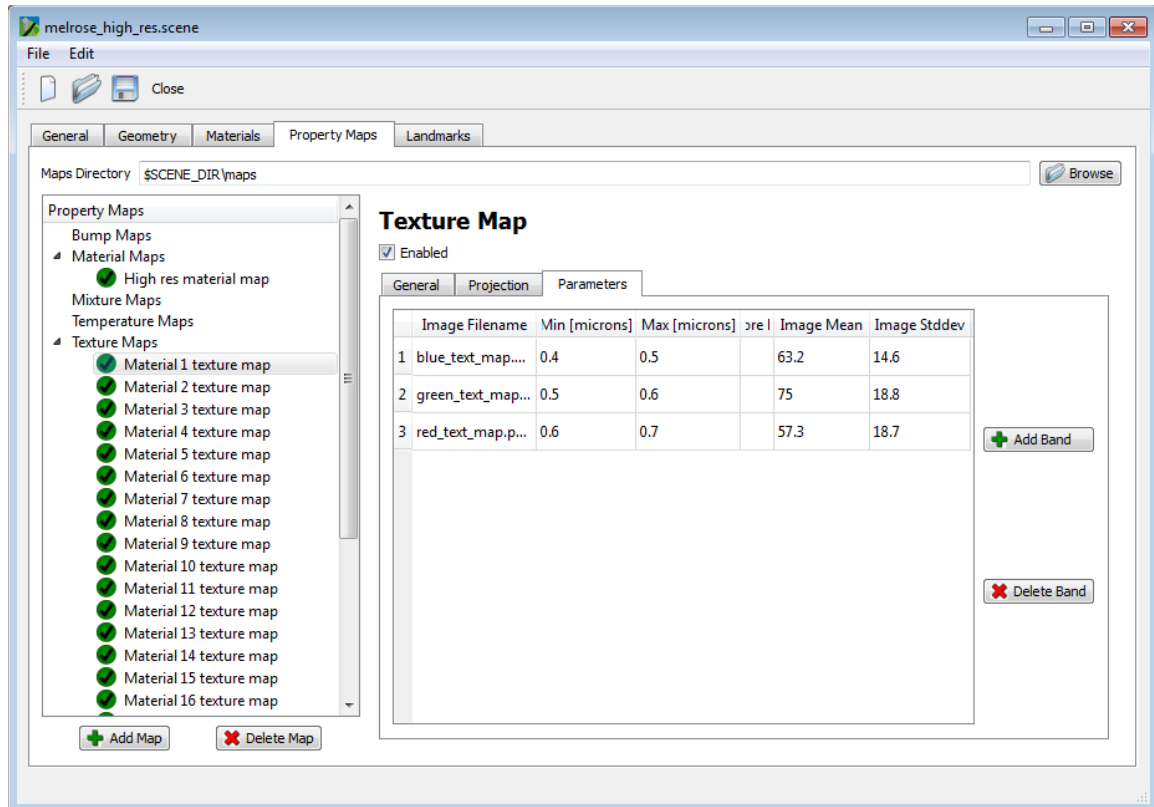
Figure 71. The Property Maps tab for the DIRSIG scene GUI showing the Parameters tab for the first texture map.

Bibliography

Auer, S., Balz, T., Becker, S., Bamler, R., 2010. 3D SAR simulation of urban areas based on detailed building models, *Photogrammetric Engineering & Remote Sensing*, Vol. 76, No. 12, December 2010, pp. 1373-1384.

Berk, A., Bernstein, L., Anderson, G., Acharya, P., Robertson, D., Chetwynd, J., Adler-Golden, S., 1998. MODTRAN cloud and multiple scattering upgrades with application to AVIRIS, *Remote Sensing of Environment*, Vol. 65, 1998, pp. 367-375.

Bolles, R., Fischler, M., 1981. A RANSAC-based approach to model fitting and its applications to finding cylinders in range data, *IJCAI*, Vol. 1981, pp. 637-643.

Brenner, C., 2005. Building reconstruction from laser scanning and images, *International Journal of Applied Earth Observation Geoinformation*, Vol. 6, No. 3, pp. 187-198.

Brenner, C., 1999. Interactive modeling tools for 3D building reconstruction, D. Fritsch and R. Spiller (eds), *Photogrammetric Week 99*, Wichmann Verlag, pp. 23–34.

Celeux, G., Diebolt, J., 1986. L'algorithme SEM: an algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densites, *Revue de Statistique Applique*, Vol. 34, No. 2, 1986.

Chang, C., Du, Q., 2004. Estimation of number of spectrally distinct signal sources in hyperspectral imagery, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, Issue 3, March 2004, pp. 608-619.

Chang, C., Plaza, A., 2006. A fast iterative algorithm for implementation of pixel purity index, *IEEE Geoscience and Remote Sensing Letters*, Vol. 3, No. 1, January 2006, pp. 63 – 67.

Cheng, L., Gong, J., Li, M., Liu, Y., 2011. 3D building model reconstruction from multi-view aerial imagery and lidar data, *Photogrammetric Engineering & Remote Sensing*, Vol. 77, No. 2, February 2011, pp. 125-139.

Eismann, M., 2012. *Hyperspectral Remote Sensing*, SPIE, Bellingham, Washington, pp. 530, 598-605.

Föerstner, W., 1999. 3D-city models: automatic and semiautomatic acquisition methods, D. Fritsch and R. Spiller (eds), *Photogrammetric Week 99*, pp. 291-303.

Givens, R., Walli, K., Eismann, M., 2012. Fusion of LIDAR Data with Hyperspectral and HR imagery for Automation of DIRSIG Scene Generation, *Proceedings of Applied*

Imagery Pattern Recognition Workshop (AIPR), IEEE, 9-11 October, 2012, Washington D.C., pp. 1-7.

Givens, R., Walli, K., Eismann, M., 2013. A method to generate sub-pixel classification maps for use in DIRSIG three-dimensional models, *Proceedings of SPIE*, Vol 8743, 18 May, 2013, Baltimore, Maryland, pp1-9.

Givens, R., Walli, K., Eismann, M., 2013. A multimodal approach to high-resolution image classification, *Proceedings of Applied Imagery Pattern Recognition Workshop (AIPR)*, IEEE, 23-25 October, 2012, Washington D.C., pp. 1-7.

Givens, R., Walli, K., Eismann, M., Schott, J., 2014. Using image fusion to generate high-resolution synthetic scenes, *Submitted to Photogrammetric Engineering & Remote Sensing*, June 2014.

Gurram, P., Rhody, H., Kerekes, J., Lach, S., Saber, E., 2007. 3D scene reconstruction through a fusion of passive video and lidar imagery, *Proceedings of the 36th Applied Imagery Pattern Recognition Workshop*, IEEE, 10-12 October, 2007, Washington D.C., pp. 133-138.

Hartley, R., Zisserman, A. 2003. *Multiple View Geometry in Computer Vision* (2nd ed.), Cambridge University Press., Cambridge, England.

Hearn, D., Baker, M., 1997. *Computer Graphics C Version*: (2nd ed), Prentice Hall, Saddle River, NJ, pp. 494-563.

Ientilucci, E., Brown, S., 2003. Advances in wide area hyperspectral image simulation, Proceedings of SPIE Aerosense, *Targets and Backgrounds IX: Characterization and Representation*, Vol. 5075, 21-25 April, 2003, Orlando, Florida, pp. 110-121.

Khoshelham, K., Li, Z., King, B., 2005. A split-and-merge technique for automated reconstruction of roof planes, *Photogrammetric Engineering & Remote Sensing*, Vol. 71, No. 7, July 2005, pp. 855-862.

Krause, K., Kuester, M., 2011. Airborne Observation Platform (AOP) Pathfinder 2010 data release, National Ecological Observatory Network, URL: http://neoninc.org/pds/files/NEON.AOP.015068.pdf, (last date accessed: May 2014).

Lach, S., Kerekes, J., Fan, X., 2009. Fusion of multiple image types for the creation of radiometrically-accurate synthetic scenes, *Journal of Applied Remote Sensing*, Vol. 3, No.1, pp. 033501-033501.

Lee, D., Shan, J., 2003. Combining lidar elevation data and ikonos multispectral imagery for coastal material mapping, *Marine Geodesy*, Vol. 26, pp. 117-127.

Li, W., Guo, Q., Jakubowski, M., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud, *Photogrammetric Engineering & Remote Sensing*, Vol. 78, No. 1, January 2012, pp. 75-84.

Lin, C., Thomson, G., Lo, C., Yang, M., 2011. A multi-level morphological active contour algorithm for delineating tree crowns in mountainous forest, *Photogrammetric Engineering & Remote Sensing*, Vol. 77, No. 3, March 2011, pp. 241-249.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110.

Masson, P., Pieczynski, W., 1993. SEM algorithm and unsupervised statistical segmentation of satellite images, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 3, May 1993, pp. 618 – 633.

Mason, T., 2004. Rapid mapping of the 3D urban environment, *Proceedings of the 20th ISPRS Congress Technical Commission III*, Volume 35, Part B3, 12-23 July 2004, Istanbul Turkey, pp. 1146-1167.

MATLAB R2010a and Image Processing Toolbox 7.0, The MathWorks, Inc., Natick, Massachusetts, United States.

Schenk, T., Csathó, B., 2002. Fusion of LIDAR and aerial imagery for a more complete surface description, *International Archives of Photogrammetry*, Remote Sensing and Spatial Information Sciences, Vol. 34, Part 3, pp. 310-317.

Schott, J., Salvaggio, C., Brown, S., Rose, R., 1995. Incorporation of texture in multispectral synthetic image generation tools, *Proceedings of SPIE Aerosense, Targets and Backgrounds: Characterization and Representation*, Vol. 2469, No. 23, April,1995, Orlando, Florida, pp. 189-196.

Schott, J., 2007. Chapter 14 Image Modeling, *Remote Sensing: The Image Chain Approach*, (2nd ed.), Oxford University Press, New York, New York (2007), pp. 617-647.

Schott, J., Gerace, A., Brown, S., Gartley, M., Montanaro, M., Reuter, D., 2012. Simulation of image performance characteristics of the landsat data continuity mission (LDCM) thermal infrared sensors (TIRS), *Remote Sensing*, Vol. 4, pp 2477-2491.

Schott, J., Brown, S., Raqueño, R., Robinson, G., 1999. An advanced synthetic image generation model and its applications to multi/hyperspectral algorithm development, *Canadian Journal of Remote Sensing*, Vol. 25, No. 2, pp. 99-111.

Visual Computing Lab ISTI-CNR, *Meshlab*, URL: http://meshlab.sourceforge.net/, (last date accessed: 13 May 2014).

Walli, K., 2003. Automated multisensory image registration. *Proceedings of Applied Imagery Pattern Recognition Workshop (AIPR)*, IEEE, 15-17 October, 2003, Washington D.C., pp. 103-107.

Walli, K., 2010. *Relating multimodal imagery data in 3D*, Ph.D. dissertation, Rochester Institute of Technology, Rochester, New York, pp. 188-189.

Walli, K., Nilosek, D., Schott, J., Salvaggio, C., 2009. Airborne synthetic scene generation (AeroSynth), *ASPRS –MAPPS Fall Conference Proceedings*, 16-19 November 2009, San Antonio, Texas.

Winter, M., 1999. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, *Proceedings of SPIE 3753, Imaging Spectroscopy V*, Vol. 266, October 27, 1999.

Wu, J., Cawse-Nicholson, K., Aardt, J., 2013. 3D tree reconstruction from simulated small footprint waveform lidar, *Photogrammetric Engineering & Remote Sensing*, Vol. 79, No. 12, December 2013, pp. 1147-1157.

Zhang, C., Qiu, F., 2012. Mapping individual tree species in an urban forest using airborne lidar data and hyperspectral imagery, *Photogrammetric Engineering & Remote Sensing*, Vol. 78, No. 10, October 2012, pp. 1079-1087.

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 074-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 08-08-2014	2. REPORT TYPE Dissertation	3. DATES COVERED *(From – To)* March 2012 – July 2014
TITLE AND SUBTITLE Automated Synthetic Scene Generation		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Givens, Ryan N., Major, USAF		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENY) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865		**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENP-DS-14-S-08
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory, Sensors Directorate 2241 Avionics Circle, Bldg 600 Wright Patterson AFB, OH 45433 DSN: 798-8506, Michael.talbert@wpafb.af.mil ATTN: Dr. Michael Talbert		**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/RYA
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
     DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Physics-based simulations generate synthetic imagery to help organizations anticipate system performance of proposed remote sensing systems. However, manually constructing synthetic scenes which are sophisticated enough to capture the complexity of real-world sites can take days to months depending on the size of the site and desired fidelity of the scene. This research successfully developed an automated approach to fuse HR imagery, lidar data, and hyperspectral imagery to extract the necessary scene components. The method greatly reduces the time and money required to generate realistic synthetic scenes and developed new approaches to improve material identification using information from all three of the input datasets.

15. SUBJECT TERMS
     Physics-based modeling, fusion, classification, synthetic imagery, synthetic scene

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Colonel Karl Walli, AFIT/ENP
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	150	19b. TELEPHONE NUMBER *(Include area code)* (703)808-4932 (wallikar@nro.mil)
U	U	U			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18